```
=> fil pascal lifesci biosis anabstr wpids scisearch
FILE 'PASCAL' ENTERED AT 16:20:07 ON 08 JAN 2004
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2004 INIST-CNRS. All rights reserved.

FILE 'LIFESCI' ENTERED AT 16:20:07 ON 08 JAN 2004
COPYRIGHT (C) 2004 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:20:07 ON 08 JAN 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'ANABSTR' ENTERED AT 16:20:07 ON 08 JAN 2004
COPYRIGHT (c) 2004 THE ROYAL SOCIETY OF CHEMISTRY (RSC)

FILE 'WPIDS' ENTERED AT 16:20:07 ON 08 JAN 2004
COPYRIGHT (C) 2004 THOMSON DERWENT

FILE 'SCISEARCH' ENTERED AT 16:20:07 ON 08 JAN 2004
COPYRIGHT 2004 THOMSON ISI

=> d que l65

L46        31499 SEA WANG Z?/AU
L47          352 SEA VOIGT C?/AU
L48          579 SEA MAYO S?/AU
L49         1553 SEA ARNOLD F?/AU
L65           13 SEA L46 AND L47 AND L48 AND L49


=> fil capl; d que l5

FILE 'CAPLUS' ENTERED AT 16:20:09 ON 08 JAN 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)


Copyright of the articles to which records in this database refer is
held by the publishers listed in the PUBLISHER (PB) field (available
for records published or updated in Chemical Abstracts after December
26, 1996), unless otherwise indicated in the original publications.
The CA Lexicon is the copyrighted intellectual property of the
American Chemical Society and is provided to assist you in searching
databases on STN.  Any dissemination, distribution, copying, or storing
of this information, without the prior written consent of CAS, is
strictly prohibited.

FILE COVERS 1907 - 8 Jan 2004  VOL 140 ISS 2
FILE LAST UPDATED: 7 Jan 2004  (20040107/ED)

  This file contains CAS Registry Numbers for easy and accurate
  substance identification.

'OBI' IS DEFAULT SEARCH FIELD FOR 'CAPLUS' FILE


L1         23639 SEA FILE=CAPLUS ABB=ON  WANG Z?/AU
L2           180 SEA FILE=CAPLUS ABB=ON  VOIGT C?/AU
L3           170 SEA FILE=CAPLUS ABB=ON  MAYO S?/AU
L4           861 SEA FILE=CAPLUS ABB=ON  ARNOLD F?/AU
```

inventors

```
L5            7 SEA FILE=CAPLUS ABB=ON  L1 AND L2 AND L3 AND L4


=> fil inspec; d que l89;d que l90

FILE 'INSPEC' ENTERED AT 16:20:09 ON 08 JAN 2004
Compiled and produced by the IEE in association with FIZ KARLSRUHE
COPYRIGHT 2004 (c) INSTITUTION OF ELECTRICAL ENGINEERS (IEE)

FILE LAST UPDATED: 2 JAN 2004       <20040102/UP>
FILE COVERS 1969 TO DATE.

<<< SIMULTANEOUS LEFT AND RIGHT TRUNCATION AVAILABLE IN
    THE BASIC INDEX >>>


L71        6780 SEA FILE=INSPEC ABB=ON  WANG Z?/AU
L72          55 SEA FILE=INSPEC ABB=ON  VOIGT C?/AU
L73          62 SEA FILE=INSPEC ABB=ON  MAYO S?/AU
L74         201 SEA FILE=INSPEC ABB=ON  ARNOLD F?/AU
L89           0 SEA FILE=INSPEC ABB=ON  L71 AND L72 AND L73 AND L74



L71        6780 SEA FILE=INSPEC ABB=ON  WANG Z?/AU
L72          55 SEA FILE=INSPEC ABB=ON  VOIGT C?/AU
L73          62 SEA FILE=INSPEC ABB=ON  MAYO S?/AU
L74         201 SEA FILE=INSPEC ABB=ON  ARNOLD F?/AU
L75       17600 SEA FILE=INSPEC ABB=ON  CROSSOVER OR CROSS? OVER
L76       47891 SEA FILE=INSPEC ABB=ON  RECOMBINATION
L90           0 SEA FILE=INSPEC ABB=ON  (L71 OR L72 OR L73 OR L74) AND L75 AND
                L76

=> dup rem l5,l65

FILE 'CAPLUS' ENTERED AT 16:20:10 ON 08 JAN 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'LIFESCI' ENTERED AT 16:20:10 ON 08 JAN 2004
COPYRIGHT (C) 2004 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:20:10 ON 08 JAN 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'WPIDS' ENTERED AT 16:20:10 ON 08 JAN 2004
COPYRIGHT (C) 2004 THOMSON DERWENT

FILE 'SCISEARCH' ENTERED AT 16:20:10 ON 08 JAN 2004
COPYRIGHT 2004 THOMSON ISI
PROCESSING COMPLETED FOR L5
PROCESSING COMPLETED FOR L65
L101          9 DUP REM L5 L65 (11 DUPLICATES REMOVED)
                ANSWERS '1-7' FROM FILE CAPLUS
                ANSWERS '8-9' FROM FILE WPIDS

=> d ibib ab 1-9

L101 ANSWER 1 OF 9  CAPLUS  COPYRIGHT 2004 ACS on STN  DUPLICATE 1
ACCESSION NUMBER:        2003:590080  CAPLUS
```

TITLE: Library analysis of SCHEMA-guided protein recombination
AUTHOR(S): Meyer, Michelle M.; Silberg, Jonathan J.; **Voigt, Christopher A.**; Endelman, Jeffrey B.; **Mayo, Stephen L.**; **Wang, Zhen-Gang**; **Arnold, Frances H.**
CORPORATE SOURCE: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125, USA
SOURCE: Protein Science (2003), 12(8), 1686-1693
CODEN: PRCIEI; ISSN: 0961-8368
PUBLISHER: Cold Spring Harbor Laboratory Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The computational algorithm SCHEMA was developed to est. the disruption caused when amino acid residues that interact in the three-dimensional structure of a protein are inherited from different parents upon recombination. To evaluate how well SCHEMA predicts disruption, we have shuffled the distantly-related .beta.-lactamases PSE-4 and TEM-1 at 13 sites to create a library of 214 (16,384) chimeras and examd. which ones retain lactamase function. Sequencing the genes from ampicillin-selected clones revealed that the percentage of functional clones decreased exponentially with increasing calcd. disruption (E = the no. of residue-residue contacts that are broken upon recombination). We also found that chimeras with low E have a higher probability of maintaining lactamase function than chimeras with the same effective level of mutation but chosen at random from the library. Thus, the simple distance metric used by SCHEMA to identify interactions and compute E allows one to predict which chimera sequences are most likely to retain their function. This approach can be used to evaluate crossover sites for recombination and to create highly mosaic, folded chimeras.
REFERENCE COUNT: 27 THERE ARE 27 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L101 ANSWER 2 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 2
ACCESSION NUMBER: 2002:478068 CAPLUS
DOCUMENT NUMBER: 137:258440
TITLE: Protein building blocks preserved by recombination
AUTHOR(S): **Voigt, Christopher A.**; Martinez, Carlos; **Wang, Zhen-Gang**; **Mayo, Stephen L.**; **Arnold, Frances H.**
CORPORATE SOURCE: Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, CA, 91125, USA
SOURCE: Nature Structural Biology (2002), 9(7), 553-558
CODEN: NSBIEW; ISSN: 1072-8368
PUBLISHER: Nature Publishing Group
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Borrowing concepts from the schema theory of genetic algorithms, we have developed a computational algorithm to identify the fragments of proteins, or schemas, that can be recombined without disturbing the integrity of the three-dimensional structure. When recombination leaves these schemas undisturbed, the hybrid proteins are more likely to be folded and functional. Crossovers found by screening libraries of several randomly shuffled proteins for functional hybrids strongly correlate with those predicted by this approach. Exptl. results from the construction of hybrids of two .beta.-lactamases that share 40% amino acid identity demonstrate a threshold in the amt. of schema disruption that the hybrid protein can tolerate. To the extent that introns function to promote recombination within proteins, natural selection would serve to bias their locations to schema boundaries.
REFERENCE COUNT: 40 THERE ARE 40 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L101 ANSWER 3 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 3
ACCESSION NUMBER: 2001:868669 CAPLUS
DOCUMENT NUMBER: 136:15894
TITLE: Methods involving identification and use of sites for gene recombination in biopolymer engineering
INVENTOR(S): **Wang, Zhen-Gang; Voigt, Christopher A.; Mayo, Stephen L.; Arnold, Frances H.**
PATENT ASSIGNEE(S): California Institute of Technology, USA
SOURCE: PCT Int. Appl., 139 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001090346	A2	20011129	WO 2001-US16831	20010523
WO 2001090346	A3	20021010		
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
EP 1283877	A2	20030219	EP 2001-937702	20010523
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR				

PRIORITY APPLN. INFO.:
US 2000-207048P P 20000523
US 2000-235960P P 20000927
US 2001-283567P P 20010413
WO 2001-US16831 W 20010523

AB The invention relates to improved methods for directed evolution of polymers, including directed evolution of nucleic acids and proteins. Specifically, the methods of the invention include anal. methods for identifying "crossover locations" in a polymer. Crossovers at these locations are less likely to disrupt desirable properties of the protein, such as stability or functionality. The invention further provides improved methods for directed evolution wherein the polymer is selectively recombined at the identified "crossover locations". Crossover disruption profiles can be used to identify preferred crossover locations. Structural domains of a biopolymer can also be identified and analyzed, and domains can be organized into schema. Schema disruption profiles can be calcd., for example based on conformational energy or interat. distances, and these can be used to identify preferred or candidate crossover locations. Computer systems for implementing anal. methods of the invention are also provided. Examples of the invention include computational calcns. of regions of .beta.-lactamase in which crossovers/in vitro recombination would disrupt protein structure, calcns. of a probability distribution for disruption of protein (sub)structures of computationally-generated recombinant mutants, and comparison of a predicted protein disruption profile with exptl. obsd. recombination crossover points.

L101 ANSWER 4 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 4
ACCESSION NUMBER: 2001:618279 CAPLUS
DOCUMENT NUMBER: 135:177723
TITLE: Computationally targeted evolutionary design

INVENTOR(S): **Voigt, Christopher; Mayo, Stephen L. ; Arnold, Frances H.; Wang, Zhen-Gang**
PATENT ASSIGNEE(S): California Institute of Technology, USA
SOURCE: PCT Int. Appl., 95 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001061344	A1	20010823	WO 2001-US5043	20010216
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
US 2001051855	A1	20011213	US 2001-795500	20010216

PRIORITY APPLN. INFO.: US 2000-183171P P 20000217

AB The invention relates to improved methods for directed evolution of polymers, including directed evolution of nucleic acids and proteins. Specifically, the methods of the invention include anal. methods for identifying "structurally tolerant" residues of a polymer. Mutations of these, structurally tolerant residues are less likely to adversely affect desirable properties of a polymer sequence. The invention further provides improved methods for directed evolution wherein the structurally tolerant residues of a polymer are selectively mutated. Computer systems for implementing anal. methods of the invention are also provided.

REFERENCE COUNT: 1 THERE ARE 1 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L101 ANSWER 5 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 5
ACCESSION NUMBER: 2001:266199 CAPLUS
DOCUMENT NUMBER: 135:15699
TITLE: Computational method to reduce the search space for directed protein evolution
AUTHOR(S): **Voigt, Christopher A.; Mayo, Stephen L.; Arnold, Frances H.; Wang, Zhen-Gang**
CORPORATE SOURCE: Biochemistry Option, Divisions of Biology and Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125, USA
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (2001), 98(7), 3778-3783
CODEN: PNASA6; ISSN: 0027-8424
PUBLISHER: National Academy of Sciences
DOCUMENT TYPE: Journal
LANGUAGE: English

AB We introduce a computational method to optimize the in vitro evolution of proteins. Simulating evolution with a simple model that statistically describes the fitness landscape, we find that beneficial mutations tend to occur at amino acid positions that are tolerant to substitutions, in the limit of small libraries and low mutation rates. We transform this observation into a design strategy by applying mean-field theory to a structure-based computational model to calc. each residue's structural tolerance. Thermostabilizing and activity-increasing mutations accumulated during the exptl. directed evolution of subtilisin E and T4

lysozyme are strongly directed to sites identified by using this computational approach. This method can be used to predict positions where mutations are likely to lead to improvement of specific protein properties.

REFERENCE COUNT: 43 THERE ARE 43 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L101 ANSWER 6 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 6
ACCESSION NUMBER: 2002:127316 CAPLUS
DOCUMENT NUMBER: 136:229009
TITLE: Computationally focusing the directed evolution of proteins
AUTHOR(S): **Voigt, Christopher A.; Mayo, Stephen L.; Arnold, Frances H.; Wang, Zhen-Gang**
CORPORATE SOURCE: Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, CA, 91125, USA
SOURCE: Journal of Cellular Biochemistry (2001), (Suppl. 37), 58-63
CODEN: JCEBD5; ISSN: 0730-2312
PUBLISHER: Wiley-Liss, Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Directed evolution has proven to be a successful strategy for the modification of enzyme properties. To date, the preferred exptl. procedure has been to apply mutations or crossovers randomly throughout the gene. With the emergence of powerful computational methods, it has become possible to develop focused combinatorial searches, guided by computer algorithms. Here, we describe several computational methods that have emerged to aid the optimization of mutant libraries, the targeting of specific residues for mutagenesis, and the design of recombination expts.
REFERENCE COUNT: 31 THERE ARE 31 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L101 ANSWER 7 OF 9 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2003:118481 CAPLUS
DOCUMENT NUMBER: 138:164697
TITLE: Algorithmic identification of regions separating structural and functional domains of biopolymers in directed evolution
INVENTOR(S): **Wang, Zhen-gang; Voigt, Christopher A.; Mayo, Stephen L.; Arnold, Frances H.**
PATENT ASSIGNEE(S): USA
SOURCE: U.S. Pat. Appl. Publ., 73 pp., Cont.-in-part of U.S. Ser. No. 863,765.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2003032059	A1	20030213	US 2001-16668	20011026
US 2002045175	A1	20020418	US 2001-863765	20010523
WO 2003055978	A2	20030710	WO 2002-US34374	20021025

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU,

TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2000-207048P P 20000523
US 2000-235960P P 20000927
US 2001-283567P P 20010413
US 2001-863765 A2 20010523
US 2001-16668 A 20011026

AB The invention relates to improved methods for directed evolution of polymers, including directed evolution of nucleic acids and proteins. Specifically, the methods of the invention include anal. methods for identifying "crossover locations" in a polymer. Crossovers are sequences sepg. structurally or functionally important domains and changes at these locations are less likely to disrupt desirable properties of the protein, such as stability or functionality. The invention further provides improved methods for directed evolution wherein the polymer is selectively recombined at the identified "crossover locations". Crossover disruption profiles can be used to identify preferred crossover locations. Structural domains of a biopolymer can also be identified and analyzed, and domains can be organized into schema. Schema disruption profiles can be calcd., for example based on conformational energy or interat. distances, and these can be used to identify preferred or candidate crossover locations. Computer systems for implementing anal. methods of the invention are also provided.

L101 ANSWER 8 OF 9 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-903494 [82] WPIDS
DOC. NO. NON-CPI: N2003-721380
DOC. NO. CPI: C2003-256957
TITLE: Selection of residues of particular polymer sequence (e.g. amino acid residues) for mutation, by obtaining level of structural tolerance for residues of particular polymer sequence, and selecting structurally tolerant residues for mutation.
DERWENT CLASS: B04 D16 T01
INVENTOR(S): **ARNOLD, F H; MAYO, S L; VOIGT, C A; WANG, Z**
PATENT ASSIGNEE(S): (CALY) CALIFORNIA INST OF TECHNOLOGY
COUNTRY COUNT: 100
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2003091835	A2	20031106	(200382)*	EN	105

RW: AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SK SL SZ TR TZ UG ZM ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2003091835	A2	WO 2002-US34342	20021025

PRIORITY APPLN. INFO: US 2001-16670 20011026
AB WO2003091835 A UPAB: 20031223

NOVELTY - Residues of a particular polymer sequence are selected for mutation, by obtaining a level of structural tolerance for residues of the particular polymer sequence; and selecting structurally tolerant residues for mutation.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for:

(1) a computer system for analyzing a polymer sequence, comprising a memory; and a processor interconnected with the memory and having software component(s) causing the processor, to execute selection of residues of particular polymer sequence for mutation;

(2) a computer program product comprising a computer readable medium having software component(s) encoded, in computer readable form; and

(3) a method for directed evolution of a polymer, comprising providing a parent polymer sequence, which has properties of interest; selecting structurally tolerant residue(s) of the parent polymer sequence for mutation; generating from the parent polymer sequence, mutant polymer sequence(s) in which the selected residue(s) is mutated; and screening the mutant sequence(s) for properties of interest.

USE - The invention is for selection of residues of particular polymer sequence comprising a sequence of amino acid residues and nucleotide residues, for mutation or for directed evolution of a polymer. The polymer comprises a polypeptide, which is one of TEM-1 and PSE-4, where the amino acid residue(s) at positions 39, 90, 99, 140, 158, 198, and 227 is substituted. The polypeptide comprises a substitution(s) consisting of Q-R and Q-N at residue 39, Q-S at residue 90, Q-R at residue 99, T-K, T-A, and T-N at residue 140, H-Y at residue 158, L-I at residue 198, and A-D at residue 227. (all claimed).

ADVANTAGE - The invention eliminates reduces the random mutagenesis of known methods, and provides a more targeted approach with improved efficiency. It is straightforward and is computationally tractable.

DESCRIPTION OF DRAWING(S) - The figure is a flow diagram illustrating a selection of residues of particular polymer sequence for mutation.

Dwg.1/10

L101 ANSWER 9 OF 9 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN

ACCESSION NUMBER: 2003-598271 [56] WPIDS
CROSS REFERENCE: 2002-122019 [16]
DOC. NO. CPI: C2003-162347
TITLE: Selecting a crossover location in a first biopolymer having a first polymer sequence, for recombination with one or more second biopolymers by identifying a particular data structure having a crossover disruption below a threshold.
DERWENT CLASS: B04 D16
INVENTOR(S): **ARNOLD, F H; MAYO, S L; VOIGT, C A; WANG, Z**
PATENT ASSIGNEE(S): (CALY) CALIFORNIA INST OF TECHNOLOGY
COUNTRY COUNT: 100
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2003055978	A2	20030710	(200356)*	EN	167

RW: AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SK SL SZ TR TZ UG ZM ZW

W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE

WO 2003055978 A2 WO 2002-US34374 ` 20021025

PRIORITY APPLN. INFO: US 2001-16668 20011026

AB WO2003055978 A UPAB: 20030903

NOVELTY - Selecting a crossover location in a first biopolymer (A), for recombination with second biopolymers comprises:

(a) identifying coupling interaction between pairs of residues in (A);

(b) generating data structures;

(c) determining a crossover disruption related to the number of coupling interactions disrupted in the crossover mutant represented by the data structure; and

(d) identifying a data structure having a crossover disruption below a threshold.

DETAILED DESCRIPTION - Selecting (M1) a crossover location in a first biopolymer having a first polymer sequence, for recombination with one or more second biopolymers, each having its own second polymer sequence comprises:

(a) identifying coupling interaction between pairs of residues in the first polymer sequence;

(b) generating a data structures, each data structure representing a crossover mutant comprising a recombination of the first and a second polymer sequence where each recombination has a different crossover location;

(c) determining, for each data structure, a crossover disruption related to the number of coupling interactions disrupted in the crossover mutant represented by the data structure; and

(d) identifying, among the data structures, a particular data structure having a crossover disruption below a threshold, where the crossover location of the crossover mutant represented by the particular data structure is the identified crossover location.

INDEPENDENT CLAIMS are also included for the following:

(1) a computer system for analyzing a polymer sequence;

(2) directed evolution of a polymer;

(3) a computer program comprising a computer readable medium having one or more software components encoded in computer readable form, where the one or more software components may be loaded into a memory of a computer system;

(4) producing hybrid polymers from two or more parent polymers;

(5) producing a library of hybrid polymers;

(6) modelling the recombination of two or more parent polymers;

(7) producing recombinant oligonucleotides from two or more parent oligonucleotides by a staggered extension process;

(8) producing recombinant oligonucleotides from two or more parent oligonucleotides by an in vitro-in vivo recombination method;

(9) producing recombinant oligonucleotides from two or more parent oligonucleotides by a PCR amplification method;

(10) producing recombinant oligonucleotides from two or more parent oligonucleotides by a family shuffling method;

(11) a beta-lactamase hybrid comprising the amino acid sequence of PSE-4, substituted in part by an amino acid sequence of TEM-1, where the substitution comprises amino acid residues 164-179, 190-216, 71-216, 71-130, or 254 and higher of PSE-4 are replaced by the corresponding amino acid residues of TEM-1; and

(12) a hybrid polymer comprising a first polypeptide recombined with at least a second polypeptide at one or more crossover locations selected according to a schema disruption threshold.

USE - The method is useful for selecting a crossover location in a first biopolymer having a first polymer sequence, for recombination with one or more second biopolymers, each having its own second polymer sequence. The methods are also useful for producing hybrid polymers from

two or more parent polymers, producing a library of hybrid polymers, modelling the recombination of two or more parent polymers, producing recombinant oligonucleotides from two or more parent oligonucleotides by a staggered extension process, an in vitro-in vivo recombination method, by a PCR amplification method, or by a family shuffling method (claimed). Dwg.1/34

```
=> fil pascal lifesci biosis anabstr wpids scisearch
FILE 'PASCAL' ENTERED AT 16:22:12 ON 08 JAN 2004
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2004 INIST-CNRS. All rights reserved.

FILE 'LIFESCI' ENTERED AT 16:22:12 ON 08 JAN 2004
COPYRIGHT (C) 2004 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:22:12 ON 08 JAN 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'ANABSTR' ENTERED AT 16:22:12 ON 08 JAN 2004
COPYRIGHT (c) 2004 THE ROYAL SOCIETY OF CHEMISTRY (RSC)

FILE 'WPIDS' ENTERED AT 16:22:12 ON 08 JAN 2004
COPYRIGHT (C) 2004 THOMSON DERWENT

FILE 'SCISEARCH' ENTERED AT 16:22:12 ON 08 JAN 2004
COPYRIGHT 2004 THOMSON ISI
```

text search

```
=> d que l66; d que l68; d que l70

L50      103779 SEA CROSSOVER OR CROSS? OVER
L51      176775 SEA RECOMBINATION
L59      241073 SEA DISRUPT?
L60     1582449 SEA COUPL?
L66           8 SEA L50 AND L51 AND L59 AND L60



L50      103779 SEA CROSSOVER OR CROSS? OVER
L51      176775 SEA RECOMBINATION
L52     1207935 SEA POLYMER# OR BIOPOLYMER#
L61       51917 SEA DATA(2A) STRUCTUR?
L62      239305 SEA DATABASE#
L63      513275 SEA ALGORITHM#
L68           3 SEA L50 AND L51 AND L52 AND (L61 OR L62 OR L63)



L50      103779 SEA CROSSOVER OR CROSS? OVER
L51      176775 SEA RECOMBINATION
L53     4035553 S PROTEIN#
L54      155693 S NUCLEIC ACID#
L55      631700 S NUCLEOTIDE#
L56     1793518 S SEQUENCE#
L57     1095809 S PEPTIDE# OR POLYPEPTIDE#                    L54 OR L55) AND
L58      140428 S OLIGONUCLEOTIDE# OR OLIGO NUCLEOTIDE#
L61       51917 S DATA(2A)STRUCTUR?
L62      239305 S DATABASE#
L63      513275 S ALGORITHM#
L102        23 (L66 OR L68 OR L70) NOT L65
```

left to see remarks & search statements

previously printed w/ inventor search

```
=> fil capl;d que l29; d que l20; d que l39; d que l40; d que l44; d que l10

FILE 'CAPLUS' ENTERED AT 16:22:17 ON 08 JAN 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
```

FILE COVERS 1907 - 8 Jan 2004 VOL 140 ISS 2
FILE LAST UPDATED: 7 Jan 2004 (20040107/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

'OBI' IS DEFAULT SEARCH FIELD FOR 'CAPLUS' FILE

L6	39229	SEA FILE=CAPLUS ABB=ON	RECOMBINATION, GENETIC/CT
L15	551	SEA FILE=CAPLUS ABB=ON	L6(L)(CROSS? OVER#/OBI)
L21	580022	SEA FILE=CAPLUS ABB=ON	GENE#/CW
L22	100363	SEA FILE=CAPLUS ABB=ON	NUCLEIC ACID#/CW
L23	977263	SEA FILE=CAPLUS ABB=ON	PROTEIN#/CW
L24	339	SEA FILE=CAPLUS ABB=ON	L15 AND (L21 OR L22 OR L23)
L25	201671	SEA FILE=CAPLUS ABB=ON	MUTANT#/OBI OR MUTAT?/OBI
L26	127	SEA FILE=CAPLUS ABB=ON	L24 AND L25
L27	46715	SEA FILE=CAPLUS ABB=ON	HYBRID#/OBI
L28	6	SEA FILE=CAPLUS ABB=ON	L26 AND L27
L29	1	SEA FILE=CAPLUS ABB=ON	STA/TI AND L28

L6	39229	SEA FILE=CAPLUS ABB=ON	RECOMBINATION, GENETIC/CT
L9	15193	SEA FILE=CAPLUS ABB=ON	DISRUPT?/OBI
L15	551	SEA FILE=CAPLUS ABB=ON	L6(L)(CROSS? OVER#/OBI)
L19	9	SEA FILE=CAPLUS ABB=ON	L15 AND L9
L20	1	SEA FILE=CAPLUS ABB=ON	BILE/TI AND L19

L6	39229	SEA FILE=CAPLUS ABB=ON	RECOMBINATION, GENETIC/CT
L7	4788	SEA FILE=CAPLUS ABB=ON	CROSSOVER?/OBI OR CROSS OVER?/OBI
L21	580022	SEA FILE=CAPLUS ABB=ON	GENE#/CW
L22	100363	SEA FILE=CAPLUS ABB=ON	NUCLEIC ACID#/CW
L23	977263	SEA FILE=CAPLUS ABB=ON	PROTEIN#/CW
L30	375164	SEA FILE=CAPLUS ABB=ON	SEQUENC?/OBI
L34	36253	SEA FILE=CAPLUS ABB=ON	ALGORITHM/CT
L38	145437	SEA FILE=CAPLUS ABB=ON	PEPTIDE#/CT OR POLYPEPTIDE#/OBI OR OLIGONUCLEOTIDE#/CT
L39	4	SEA FILE=CAPLUS ABB=ON	L6 AND L7 AND (L21 OR ((L22 OR L23) AND L30) OR L38) AND L34

L6	39229	SEA FILE=CAPLUS ABB=ON	RECOMBINATION, GENETIC/CT
L7	4788	SEA FILE=CAPLUS ABB=ON	CROSSOVER?/OBI OR CROSS OVER?/OBI
L11	1126408	SEA FILE=CAPLUS ABB=ON	POLYMER#/OBI OR BIOPOLYMER#/OBI
L34	36253	SEA FILE=CAPLUS ABB=ON	ALGORITHM/CT
L40	1	SEA FILE=CAPLUS ABB=ON	L6 AND L7 AND L11 AND L34

```
L6        39229 SEA FILE=CAPLUS ABB=ON  RECOMBINATION, GENETIC/CT
L7         4788 SEA FILE=CAPLUS ABB=ON  CROSSOVER?/OBI OR CROSS OVER?/OBI
L12      123973 SEA FILE=CAPLUS ABB=ON  DATA/OBI
L13       15285 SEA FILE=CAPLUS ABB=ON  DATABASE#/OBI
L25      201671 SEA FILE=CAPLUS ABB=ON  MUTANT#/OBI OR MUTAT?/OBI
L27       46715 SEA FILE=CAPLUS ABB=ON  HYBRID#/OBI
L34       36253 SEA FILE=CAPLUS ABB=ON  ALGORITHM/CT
L43        8770 SEA FILE=CAPLUS ABB=ON  MUTAGENESIS/CT
L44           5 SEA FILE=CAPLUS ABB=ON  L6 AND L7 AND (L25 OR L27 OR L43) AND
                (L34 OR (L12 OR L13))


L6        39229 SEA FILE=CAPLUS ABB=ON  RECOMBINATION, GENETIC/CT
L7         4788 SEA FILE=CAPLUS ABB=ON  CROSSOVER?/OBI OR CROSS OVER?/OBI
L8       257749 SEA FILE=CAPLUS ABB=ON  COUPL?/OBI
L9        15193 SEA FILE=CAPLUS ABB=ON  DISRUPT?/OBI
L10           0 SEA FILE=CAPLUS ABB=ON  L6 AND L7 AND L8 AND L9


=> s (l29 or l20 or l39 or l40 or l44) not l5

L103          8 (L29 OR L20 OR L39 OR L40 OR L44) NOT (L5)  previously printed

=> fil inspec; d que l94; d que l99; d que l100

FILE 'INSPEC' ENTERED AT 16:22:19 ON 08 JAN 2004
Compiled and produced by the IEE in association with FIZ KARLSRUHE
COPYRIGHT 2004 (c) INSTITUTION OF ELECTRICAL ENGINEERS (IEE)

FILE LAST UPDATED: 2 JAN 2004        <20040102/UP>
FILE COVERS 1969 TO DATE.

<<< SIMULTANEOUS LEFT AND RIGHT TRUNCATION AVAILABLE IN
    THE BASIC INDEX >>>


L75       17600 SEA FILE=INSPEC ABB=ON  CROSSOVER OR CROSS? OVER
L76       47891 SEA FILE=INSPEC ABB=ON  RECOMBINATION
L77      154309 SEA FILE=INSPEC ABB=ON  POLYMER# OR BIOPOLYMER#
L78       33956 SEA FILE=INSPEC ABB=ON  PROTEIN#
L79        1750 SEA FILE=INSPEC ABB=ON  NUCLEIC ACID#
L80        1687 SEA FILE=INSPEC ABB=ON  NUCLEOTIDE#
L81      161769 SEA FILE=INSPEC ABB=ON  SEQUENCE#
L82        5165 SEA FILE=INSPEC ABB=ON  PEPTIDE# OR POLYPEPTIDE#
L83         608 SEA FILE=INSPEC ABB=ON  OLIGONUCLEOTIDE# OR OLIGO NUCLEOTIDE#
L86       43277 SEA FILE=INSPEC ABB=ON  DATA(2A)STRUCTUR?
L87      124443 SEA FILE=INSPEC ABB=ON  DATABASE#
L88      476863 SEA FILE=INSPEC ABB=ON  ALGORITHM#
L94           1 SEA FILE=INSPEC ABB=ON  L75 AND L76 AND (L77 OR ((L78 OR L79
                OR L80) AND L81) OR (L82 OR L83)) AND (L86 OR L87 OR L88)


L75       17600 SEA FILE=INSPEC ABB=ON  CROSSOVER OR CROSS? OVER
L76       47891 SEA FILE=INSPEC ABB=ON  RECOMBINATION
L84       12875 SEA FILE=INSPEC ABB=ON  DISRUPT?
L85      362360 SEA FILE=INSPEC ABB=ON  COUPL?
L96          20 SEA FILE=INSPEC ABB=ON  L75 AND L76 AND (L84 OR L85)
L98        7224 SEA FILE=INSPEC ABB=ON  MUTANT# OR MUTAT? OR MUTAGENESIS
```

```
L99           1 SEA FILE=INSPEC ABB=ON  L96 AND L98


L75       17600 SEA FILE=INSPEC ABB=ON  CROSSOVER OR CROSS? OVER
L76       47891 SEA FILE=INSPEC ABB=ON  RECOMBINATION
L77      154309 SEA FILE=INSPEC ABB=ON  POLYMER# OR BIOPOLYMER#
L78       33956 SEA FILE=INSPEC ABB=ON  PROTEIN#
L79        1750 SEA FILE=INSPEC ABB=ON  NUCLEIC ACID#
L80        1687 SEA FILE=INSPEC ABB=ON  NUCLEOTIDE#
L81      161769 SEA FILE=INSPEC ABB=ON  SEQUENCE#
L82        5165 SEA FILE=INSPEC ABB=ON  PEPTIDE# OR POLYPEPTIDE#
L83         608 SEA FILE=INSPEC ABB=ON  OLIGONUCLEOTIDE# OR OLIGO NUCLEOTIDE#
L84       12875 SEA FILE=INSPEC ABB=ON  DISRUPT?
L85      362360 SEA FILE=INSPEC ABB=ON  COUPL?
L86       43277 SEA FILE=INSPEC ABB=ON  DATA(2A)STRUCTUR?
L87      124443 SEA FILE=INSPEC ABB=ON  DATABASE#
L88      476863 SEA FILE=INSPEC ABB=ON  ALGORITHM#
L98        7224 SEA FILE=INSPEC ABB=ON  MUTANT# OR MUTAT? OR MUTAGENESIS
L100          9 SEA FILE=INSPEC ABB=ON  L75 AND L76 AND (L84 OR L85) AND ((L86
                OR L87 OR L88) OR (L77 OR L78 OR L79 OR L80 OR L81 OR L82 OR
                L83) OR L98)


=> s (l94 or l99 or l100)

L104         10 (L94 OR L99 OR L100)

=> dup rem l103,l104,l102
FILE 'CAPLUS' ENTERED AT 16:22:34 ON 08 JAN 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'INSPEC' ENTERED AT 16:22:34 ON 08 JAN 2004
Compiled and produced by the IEE in association with FIZ KARLSRUHE
COPYRIGHT 2004 (c) INSTITUTION OF ELECTRICAL ENGINEERS (IEE)

FILE 'PASCAL' ENTERED AT 16:22:34 ON 08 JAN 2004
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2004 INIST-CNRS. All rights reserved.

FILE 'LIFESCI' ENTERED AT 16:22:34 ON 08 JAN 2004
COPYRIGHT (C) 2004 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:22:34 ON 08 JAN 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'ANABSTR' ENTERED AT 16:22:34 ON 08 JAN 2004
COPYRIGHT (c) 2004 THE ROYAL SOCIETY OF CHEMISTRY (RSC)

FILE 'WPIDS' ENTERED AT 16:22:34 ON 08 JAN 2004
COPYRIGHT (C) 2004 THOMSON DERWENT

FILE 'SCISEARCH' ENTERED AT 16:22:34 ON 08 JAN 2004
COPYRIGHT 2004 THOMSON ISI
PROCESSING COMPLETED FOR L103
PROCESSING COMPLETED FOR L104
PROCESSING COMPLETED FOR L102
L105          33 DUP REM L103 L104 L102 (8 DUPLICATES REMOVED)
```

ANSWERS '1-8' FROM FILE CAPLUS
ANSWERS '9-18' FROM FILE INSPEC
ANSWER '19' FROM FILE PASCAL
ANSWERS '20-22' FROM FILE LIFESCI
ANSWERS '23-24' FROM FILE BIOSIS
ANSWER '25' FROM FILE ANABSTR
ANSWERS '26-27' FROM FILE WPIDS
ANSWERS '28-33' FROM FILE SCISEARCH

=> d ibib ab 1-33; fil hom

L105 ANSWER 1 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 1
ACCESSION NUMBER: 2003:757819 CAPLUS
DOCUMENT NUMBER: 139:272045
TITLE: Optimization of **crossover** points for biomolecule directed evolution
INVENTOR(S): Govindarajan, Sridhar; Gustafsson, Claes; Minshull, Jeremy S.
PATENT ASSIGNEE(S): Maxygen, Inc., USA; Mundorff, Emily C.
SOURCE: PCT Int. Appl., 78 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003078583	A2	20030925	WO 2003-US7610	20030310
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG				
US 2003198988	A1	20031023	US 2003-386903	20030310

PRIORITY APPLN. INFO.: US 2002-363505P P 20020309
US 2002-373591P P 20020418

AB Methods and devices for more efficiently engineering diversity into recombinant polypeptides and/or nucleic acids are provided herein. For example, a variety of methods of selecting and/or assessing potential crossover sites in an amino acid sequence or a nucleotide sequence are provided, as well as the resulting chimeric product sequences. These methods include, e.g., consideration of structural, functional and/or statistical data in the selection and assessment of sequences and crossover sites for use in recombination. Wild type genes for muconate lactonizing enzyme, MLE 1 and 2 from Pseudomonas putida and MLE 1 gene from Acetobacter calcoaceticus were used in examples for engyneering enzymes with composite active sites.

L105 ANSWER 2 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN DUPLICATE 3
ACCESSION NUMBER: 2002:555695 CAPLUS
DOCUMENT NUMBER: 137:93361
TITLE: A modeling framework for predicting the number, type, and distribution of **crossovers** in directed evolution experiments
INVENTOR(S): Maranas, Costas D.; Moore, Gregory
PATENT ASSIGNEE(S): The Penn State Research Foundation, USA

SOURCE: PCT Int. Appl., 50 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002057495	A2	20020725	WO 2001-US50372	20011109
WO 2002057495	A3	20031016		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG				
US 2003073092	A1	20030417	US 2001-37572	20011109

PRIORITY APPLN. INFO.:
US 2000-247088P P 20001110
US 2000-255580P P 20001214
US 2001-270362P P 20010220
US 2001-316683P P 20010831

AB A modeling framework for predicting the no., type and distribution of crossovers in directed evolution expts. is disclosed. The framework provides for detg. how fragmentation length, annealing temp., sequence identity, and no. of shuffled parent sequences affect the no., type, and distribution of crossovers along the length of reassembled sequences. This framework allows for the optimization of directed evolution protocols in response to a particular enzyme or protein design challenge. One method according to the present invention includes applying equil. thermodn. to a plurality of sequences to det. statistics of hybridization; and parameterizing an assembly algorithm using the statistics of hybridization. According to the framework of the present invention, the annealing events during reassembly are modeled as a network of reactions, and equil. thermodn. is used to quantify their conversions and selectivities. The key idea of the reassembly algorithm is to postulate a set of recursive relations that describe the probability that a full-length reassembled sequence involves a given no. of crossovers. An in silico case study of a set of 12 subtilases examines the effect of fragmentation length, annealing temp., sequence identity and no. of shuffled sequences on the no., type, and distribution of crossovers. A computational verification of crossover aggregation in regions of near-perfect sequence identity and the presence of synergistic reassembly in family DNA shuffling is obtained.

L105 ANSWER 3 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2003:482978 CAPLUS
DOCUMENT NUMBER: 139:160536
TITLE: General Method for **Sequence**-independent Site-directed Chimeragenesis
AUTHOR(S): Hiraga, Kaori; Arnold, Frances H.
CORPORATE SOURCE: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125, USA
SOURCE: Journal of Molecular Biology (2003), 330(2), 287-296
CODEN: JMOBAK; ISSN: 0022-2836
PUBLISHER: Elsevier Science Ltd.
DOCUMENT TYPE: Journal
LANGUAGE: English

AB We have developed a simple and general method that allows for the facile

recombination of distantly related (or unrelated) proteins at multiple discrete sites. To evaluate the sequence-independent site-directed chimeragenesis (SISDC) method, we have recombined .beta.-lactamases TEM-1 and PSE-4 at seven sites, examd. the quality of the chimeric genes created, and screened the library of 28 (256) chimeras for functional enzymes. Probe hybridization and sequencing analyses revealed that SISDC generated a random library with little sequence bias and in which all targeted fragments were recombined in the desired order. Sequencing the genes from clones having functional lactamases identified 14 unique chimeras. These chimeras are characterized by a lower level of disruption, as calcd. by the SCHEMA algorithm, than the library as a whole. These results illustrate the use of SISDC in creating designed chimeric protein libraries and further illustrate the ability of SCHEMA to identify chimeras whose folded structures are likely not to be disrupted by recombination.

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L105 ANSWER 4 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2002:897401 CAPLUS
DOCUMENT NUMBER: 138:164663
TITLE: Distribution of recombination **crossovers** and the origin of haplotype blocks: the interplay of population history, recombination, and **mutation**
AUTHOR(S): Wang, Ning; Akey, Joshua M.; Zhang, Kun; Chakraborty, Ranajit; Jin, Li
CORPORATE SOURCE: Center for Genome Information, University of Cincinnati, Cincinnati, OH, 45267-0056, USA
SOURCE: American Journal of Human Genetics (2002), 71(5), 1227-1234
CODEN: AJHGAG; ISSN: 0002-9297
PUBLISHER: University of Chicago Press
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Recent studies suggest that haplotypes are arranged into discrete blocklike structures throughout the human genome. Here, we present an alternative haplotype block definition that assumes no recombination within each block but allows for recombination between blocks, and we use it to study the combined effects of demog. history and various population genetic parameters on haplotype block characteristics. Through extensive coalescent simulations and anal. of published haplotype data on chromosome 21, we find that (1) the combined effects of population demog. history, recombination, and mutation dictate haplotype block characteristics and (2) haplotype blocks can arise in the absence of recombination hot spots. Finally, we provide practical guidelines for designing and interpreting studies investigating haplotype block structure.

REFERENCE COUNT: 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L105 ANSWER 5 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1997:305632 CAPLUS
DOCUMENT NUMBER: 127:31815
TITLE: Compound heterozygosity for Hb S and the **hybrid** Hbs Lepore, P-Nilotic, and Kenya; comparison of hematological and hemoglobin composition **data**
AUTHOR(S): Huisman, T.H.J.
CORPORATE SOURCE: Medical College of Georgia, Augusta, GA, 30912-2114, USA
SOURCE: Hemoglobin (1997), 21(3), 249-257
CODEN: HEMOD8; ISSN: 0363-0269
PUBLISHER: Dekker

Searched by Barb O'Bryen, STIC 308-4291

DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 31 refs. on the anal. of hematol. and Hb compn. data for patients with a heterozygosity for Hb Lepore, Hb P-Nilotic, and Hb Kenya, with and without a Hb S heterozygosity. The locations of the areas of crossover between the .delta.- and .beta.-globin genes leading to the formation of the .delta..beta. genes of Hb Lepore anomalies, and the .beta..delta. gene of the Hb P-Nilotic abnormality are discussed.
REFERENCE COUNT: 31 THERE ARE 31 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L105 ANSWER 6 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1996:54509 CAPLUS
DOCUMENT NUMBER: 124:107928
TITLE: Construction of phylogenetic trees from amino acid **sequences** using a genetic algorithm
AUTHOR(S): Matsuda, Hideo
CORPORATE SOURCE: Dep. of Information and Computer Sciences, Osaka Univ., Osaka, 560, Japan
SOURCE: Genome Informatics Series (1995), 6(Genome Informatics Workshop 1995), 19-28
CODEN: GINSE9; ISSN: 0919-9454
PUBLISHER: Universal Academy Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB We have developed a novel algorithm to search for the max. likelihood tree constructed from amino acid sequences. This algorithm is a variant of genetic algorithms which uses scores derived from the log-likelihood of trees computed by the max. likelihood method. This algorithm is valuable since it may construct a more likely tree from randomly generated trees by utilizing crossover and mutation operators. In a test of our algorithm on a data set of elongation factor-1 .alpha. sequences, we found that the performance of our algorithm is comparable to that of other tree-construction methods (UPGMA, the neighbor-joining and the max. parsimony methods; and the max. likelihood method with different search algorithms).

L105 ANSWER 7 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1993:532469 CAPLUS
DOCUMENT NUMBER: 119:132469
TITLE: Gene **disruption** in Lactobacillus plantarum strain 80 by site-specific recombination: Isolation of a mutant strain deficient in conjugated **bile** salt hydrolase activity
AUTHOR(S): Leer, R. J.; Christiaens, H.; Verstraete, W.; Peters, L.; Posno, M.; Pouwels, P. H.
CORPORATE SOURCE: Med. Biol. Lab., TNO, Rijswijk, 5815 HV, Neth.
SOURCE: Molecular and General Genetics (1993), 239(1-2), 269-72
CODEN: MGGEAE; ISSN: 0026-8925
DOCUMENT TYPE: Journal
LANGUAGE: English
AB A chloramphenicol-resistance gene (cml) was introduced into the Lactobacillus plantarum gene encoding conjugated bile acid hydrolase (cbh) on a ColE1 replicon. This plasmid which is nonreplicative in Lactobacillus was used to transform L. plantarum strain 80. A homologous double cross-over recombination event resulted in replacement of the chromosomal cbh gene by the cml-contg. cbh gene. The transformants obtained were unable to synthesize active conjugated bile acid hydrolase (Cbh). The Cbh- CmlR phenotype was stably maintained for more than 100 generations under nonselective conditions.

L105 ANSWER 8 OF 33 CAPLUS COPYRIGHT 2004 ACS on STN

ACCESSION NUMBER: 1990:453523 CAPLUS
DOCUMENT NUMBER: 113:53523
TITLE: Identification of the crossing-over point of a **hybrid** gene encoding human glycophorin variant **Sta**. Similarity to the crossing-over point in haptoglobin-related genes
AUTHOR(S): Rearden, Ann; Phan Huan; Dubnicoff, Todd; Kudo, Shinichi; Fukuda, Minoru
CORPORATE SOURCE: Dep. Pathol., Univ. California, San Diego, La Jolla, CA, 92093, USA
SOURCE: Journal of Biological Chemistry (1990), 265(16), 9259-63
CODEN: JBCHA3; ISSN: 0021-9258
DOCUMENT TYPE: Journal
LANGUAGE: English

AB One of the human glycophorin variants, Stones (Sta), has been shown to be the product of a hybrid gene of which the 5'-half is derived from the glycophorin B (GPB) gene and the 3'-half is derived from the glycophorin A (GPA) gene. The present study reveals the crossing-over point of this hybrid gene from the anal. of polymerase chain reaction products. The genomic sequences encompassing the region corresponding to exon 3 to exon 4 of GPA were amplified by polymerase chain reaction with oligonucleotide primers synthesized according to GPA and GPB genomic sequences. After subcloning the products, the nucleotide sequences derived from GPA, GPB, and putative Sta genes were detd. Comparison of the nucleotide sequences of GPA, GPB, and Sta genes indicate that the crossing-over took place 200 bp upstream from the first nucleotide of exon 4. Intriguingly, the nucleotide sequence surrounding the putative crossing-over point is homologous to the crossing-over point proposed for haptoglobin genes (Maeda, N., et al., 1986). These results suggest strongly that homologous recombination through unequal crossing-over can be facilitated by specific genomic elements, such as those in common in these 2 crossing-over events. The present study also revealed that this Sta individual has a variant GPA gene; substitution of adenine for guanine at the nucleotide for codon 39 results in substitution of lysine for arginine at amino acid 39, and loss of an SstI restriction site.

L105 ANSWER 9 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 2002:7380885 INSPEC
DOCUMENT NUMBER: A2002-21-8715-002; B2002-10-2230B-006
TITLE: Long-range and very long-range charge transport in DNA.
AUTHOR: Bixon, M.; Jortner, J. (Sch. of Chem., Tel Aviv Univ., Israel)
SOURCE: Chemical Physics (1 Aug. 2002) vol.281, no.2-3, p.393-408. 65 refs.
Doc. No.: S0301-0104(02)00495-0
Published by: Elsevier
Price: CCCC 0301-0104/02/$22.00
CODEN: CMPHC2 ISSN: 0301-0104
SICI: 0301-0104(20020801)281:2/3L.393:LRVL;1-S
DOCUMENT TYPE: Journal
TREATMENT CODE: Theoretical
COUNTRY: Netherlands
LANGUAGE: English

AB We present a kinetic-quantum model for the mechanisms of hole transport in DNA duplexes, which involves a **sequence** of hole hopping processes between adjacent guanines (G) and/or hole hopping/trapping via GG or GGG, all of which are separated by thymine (T)-adenine (A) bridges. Individual hole hopping processes between G sites fall into two distinct parallel mechanisms, i.e., unistep superexchange mediated hopping via 'short' (T-A), bridges and thermally induced hopping (TIH) via 'long' (T-A)n (n>3-4) bridges. The bridge specificity for TIH via (A)n chains

pertains to the energetics, with the G+A energy gap Delta =0.20+or-0.05 eV being sufficiently low to warrant endothermic hole excitation from G+ to (A)n, and to the electronic **couplings**, with the nearest-neighbor A-A **couplings** being unique in the sense that the intrastrand and interstrand **couplings** are close and large (V(A-A) approximately=0.30-0.060 eV). Accordingly, both effective intrastrand and interstrand (zigzagging) hole transport via (A), chains will prevail, being nearly invariant with respect to the nucleobases ordering within the (T-A)n duplex. We treated the 'transition' between the superexchange and the TIH mechanism in 5'-G+(T-A)nG-3' duplexes to predict that the **crossover** occurs at nx approximately=3-4, with nx exhibiting a moderate bridge specificity and energy gap dependence. nx is in accord with the experimental data of Giese et al. [Nature 412, 318, 2001]. We assert that the kinetic-quantum mechanical model for the chemical yields and elementary rates cannot be reconciled with the experimental TIH data, with respect to the very weak bridge size dependence of the relative chemical yields and the ratios of the rates. Configurational relaxation accompanying endothermic hole injection from G+ to (A)n may result in the gating (switching-off) of the backrecombination, providing a reasonable description of TIH dynamics and very long-range hole transport in long (A)n chains.

L105 ANSWER 10 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 2002:7310386 INSPEC
DOCUMENT NUMBER: C2002-08-1290L-020
TITLE: The immune and the chemical **crossover**.
AUTHOR: Bersini, H. (IRIDIA Lab., Univ. Libre de Bruxelles, Brussels, Belgium)
SOURCE: IEEE Transactions on Evolutionary Computation (June 2002) vol.6, no.3, p.306-13. 30 refs.
Doc. No.: S1089-778X(02)06068-X
Published by: IEEE
Price: CCCC 1089-778X/02/$17.00
CODEN: ITEVF5 ISSN: 1089-778X
SICI: 1089-778X(200206)6:3L.306:ICC;1-6
DOCUMENT TYPE: Journal
TREATMENT CODE: Theoretical
COUNTRY: United States
LANGUAGE: English
AB Among the different mechanisms employed by evolutionary **algorithms**, it can be argued that **recombination**, or **crossover**, is the most original, intuitively appealing and useful in an engineering perspective. It is a simple, but natural trick to combine elements of two good individuals in the hopes of generating a better one and, in particular, by combining the elements that make these solutions good in isolation. The trick of **recombination** can be seen not only in genetic systems, but also in immune and chemical systems as well. This paper describes and explains these latter **recombination** mechanisms, first from a biological or chemical perspective, then from an engineering perspective. With regard to **crossover** in immune systems, several algorithmic mechanisms have already been proposed (e.g. IRM, GA-Simplex, STEP) and these are reviewed. Their basic functionality in each case is the same: new individuals are created in a zone of the search space that is shaped by the position of the current solutions, together with their fitness values. When the immune system proposes a new cell, the profile of this new candidate evidences a huge diversity, providing its adaptive capability, but this is subject to a subsequent "recruitment test" under the selective pressure of the current population of cells. With regard to **crossover** in chemical reactions, these can be viewed as a combination of computational graphs **coupled** with the distribution of the fitness values assigned to components in the graphs, as is already evidenced in particular instances of genetic **algorithms** and genetic programming. The benefits that these new

features allow are discussed, along with other possible positive influences that come from chemistry. Finally, the paper shows how chemistry and immunology converge to this same basic message, which is in line with classical optimization techniques: exploit the information contained in the current population of solutions better before proposing a new candidate to be evaluated.

L105 ANSWER 11 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 2001:7010689 INSPEC
DOCUMENT NUMBER: A2001-18-7145L-002
TITLE: Confinement, dimensional **crossover** and topological **coupling** in quasi one dimensional electronic systems.
AUTHOR: Brazovskii, S. (LPTMS, Univ. de Paris-Sud, Orsay, France)
SOURCE: Synthetic Metals (15 March 2001) vol.120, no.1-3, p.691-4. 12 refs.
Doc. No.: S0379-6779(00)01128-0
Published by: Elsevier
Price: CCCC 0379-6779/2001/$20.00
CODEN: SYMEDZ ISSN: 0379-6779
SICI: 0379-6779(20010315)120:1/3L.691:CDCT;1-N
Conference: International Conference on Science and Technology of Synthetic Metals. Gastein, Austria, 15-21 July 2000
DOCUMENT TYPE: Conference Article; Journal
TREATMENT CODE: Theoretical
COUNTRY: Switzerland
LANGUAGE: English
AB Topologically nontrivial states are common in symmetry broken phases at macroscopic scales. Low dimensional systems bring them to a microscopic level where solitons emerge as single particles. The examples are conducting **polymers** and spin-Peierls chains. We shall discuss topological aspects of elementary excitations, especially the confinement and the dimensional D **crossover**. At D>1 the topological requirements for the combined symmetry originate the spin- or charge-roton like excitations with charge- or spin- kinks localized in the core. In a quasi 1D world they can be viewed as resulting from a spin-charge **recombination** due to the 2D or 3D confinement.

L105 ANSWER 12 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 2001:7128735 INSPEC
DOCUMENT NUMBER: C2002-01-6150N-128
TITLE: Conventional and multirecombinative evolutionary **algorithms** for the parallel task scheduling problem.
AUTHOR: Esquivel, S.; Gatica, C.; Gallard, R. (Laboratorio de Investigacion y Desarrollo en Inteligencia Computacional, Univ. Nacional de San Luis, Argentina)
SOURCE: Applications of Evolutionary Computing. EvoWorkshops 2001: EvoCOP, EvoFlight, EvoIASP, EvoLearn, and EvoSTIM. Proceedings (Lecture Notes in Computer Science Vol.2037)
Editor(s): Boers, E.J.W.;
Berlin, Germany: Springer-Verlag, 2001. p.223-32 of xiv+516 pp. 24 refs.
Conference: Como, Italy, 18-20 April 2001
Sponsor(s): EvoNet
ISBN: 3-540-41920-9
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Application; Practical
COUNTRY: Germany, Federal Republic of
LANGUAGE: English

Searched by Barb O'Bryen, STIC 308-4291

AB This paper deals with the problem of allocating a number of non identical tasks in a parallel system. The model assumes that the system consists of a number of identical processors and that only one task may be executed on a processor at a time. All schedules and tasks are nonpreemptive. R.L. Graham's (1972) well-known list scheduling **algorithm** (LSA) is contrasted with different evolutionary **algorithms** (EAs), which differ on the representations and the recombinative approach used. Regarding representation, direct and indirect representation of schedules are used. Concerning **recombination**, the conventional single **crossover** per **couple** (SCPC) and a multiple **crossover** per **couple** (MCPC) are used. Outstanding behaviour of evolutionary **algorithms** when contrasted against LSA was detected. Results are shown and discussed.

L105 ANSWER 13 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1999:6338995 INSPEC
DOCUMENT NUMBER: C1999-10-1180-056
TITLE: Multiple crossovers between multiple parents to improve search in evolutionary **algorithms**.
AUTHOR: Esquivel, S.C. (Dept. de Inf., Univ. Nacional de San Luis, Argentina); Leiva, H.A.; Gallard, R.H.
SOURCE: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)
Piscataway, NJ, USA: IEEE, 1999. p.1589-94 Vol. 2 of 3 vol. (xxxvii+2348) pp. 19 refs.
Conference: Washington, DC, USA, 6-9 July 1999
Price: CCCC 0 7803 5536 9/99/$10.00
ISBN: 0-7803-5536-9
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Theoretical
COUNTRY: United States
LANGUAGE: English
AB As a new promising **crossover** method, multiple crossovers per **couple** (MCPC) deserves special attention in the evolutionary computing field. Allowing multiple crossovers per **couple** on a selected pair of parents provided an extra benefit in processing time and similar quality of solutions when contrasted against the conventional approach, which applies a single **crossover** operation per **couple**. These results were confirmed when optimising classic testing functions and harder (non-linear, non-separable) functions. Despite these benefits, due to a reinforcement of selective pressure, MCPC showed in some cases an undesirable premature convergence effect. An adequate balance between exploitation and exploration can improve search. Extreme exploitation can lead to premature convergence and intense exploration can make the search ineffective. Focussing on this equilibrium problem, a previous proposal combined MCPC with an alternative selection method; fitness proportional **couple** selection (FPCS) which first creates an intermediate population **couples** where both individuals were chosen proportional selection. Then a criterion is applied to establish the fitness of a **couple** and subsequently, **couples** are selected for **crossing-over** based on **couple** fitness. This paper investigates the raw effect in performance on a pair of selected optimization problems by using a new multiple crossovers on multiple parents (MCMP) method, which allows multiple **recombination** of multiple parents under uniform scanning **crossover**.

L105 ANSWER 14 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1997:5629846 INSPEC
DOCUMENT NUMBER: C9708-6110P-030
TITLE: Preserving locality for optimal parallelism in task allocation.
AUTHOR: Schoneveld, A.; De Ronde, J.F.; Sloot, P.M.A. (Dept.

of Math., Amsterdam Univ., Netherlands)
SOURCE: High-Performance Computing and Networking. International Conference and Exhibition. Proceedings
Editor(s): Hertzberger, B.; Sloot, P.
Berlin, Germany: Springer-Verlag, 1997. p.565-74 of xxi+1066 pp. 13 refs.
Conference: Vienna, Austria, 28-30 April 1997
ISBN: 3-540-62898-3
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Practical; Theoretical; Experimental
COUNTRY: Germany, Federal Republic of
LANGUAGE: English
AB Genetic **algorithms** have been applied to several combinatorial optimisation problems, including the well-known task allocation problem, originating from parallel computing. We introduce random task graphs as a model of applications which display irregular global communication patterns. Uniform **crossover** is the standard genetic **recombination** operator that is applied to solution-encoded chromosomes. However, application of a uniform **crossover** may heavily **disrupt** low-cost sub-solutions, or building blocks, of a chromosome. Therefore, we define a locality-preserving **recombination** operator, exploiting the connectivity of the task graph. Experiments show that this new operator increases the convergence rate of the genetic **algorithm** applied to the task allocation problem.

L105 ANSWER 15 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1995:5080342 INSPEC
DOCUMENT NUMBER: B9511-0260-031; C9511-1180-074
TITLE: The usefulness of **recombination**.
AUTHOR: Hordijk, W.; Manderick, B. (Sante Fe Inst., NM, USA)
SOURCE: Advances in Artificial Life. Third European Conference on Artificial Life Proceedings
Editor(s): Moran, F.; Moreno, A.; Merelo, J.J.; Chacon, P.
Berlin, Germany: Springer-Verlag, 1995. p.908-19 of xiii+960 pp. 6 refs.
Conference: Granada, Spain, 4-6 June 1995
ISBN: 3-540-59496-5
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Theoretical
COUNTRY: Germany, Federal Republic of
LANGUAGE: English
AB In this paper, we examine the usefulness of **recombination** from two points of view. First, the problem of **crossover disruption** is investigated. This is done by comparing two genetic **algorithms** with different **crossover** operators (one-point and uniform) to each other on NK-landscapes with different values of K relative to N, and with different epistatic interactions (random and nearest neighbor). Second, the usefulness of **recombination** in relation to the location of local optima in the fitness landscape is investigated. There appears to be a clear relation between the type of fitness landscape and the type of **recombination** that is most useful on this landscape. Furthermore, there also is a clear relation between the location of local optima in the fitness landscape and the usefulness of **recombination**.

L105 ANSWER 16 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1996:5248095 INSPEC
DOCUMENT NUMBER: C9606-1180-058
TITLE: Raising GA performance by simultaneous tuning of selective pressure and **recombination disruptiveness**.

Searched by Barb O'Bryen, STIC 308-4291

AUTHOR: Van Kemenade, C.H.M. (CWI, Amsterdam, Netherlands); Kok, J.N.; Eiben, A.E.
SOURCE: 1995 IEEE International Conference on Evolutionary Computation (Cat. No.95TH8099)
New York, NY, USA: IEEE, 1995. p.346-51 vol.1 of 2 vol. xii+855 pp. 9 refs.
Conference: Perth, WA, Australia, 29 Nov-1 Dec 1995
Sponsor(s): IEEE Neural Network Council
Price: CCCC 0 7803 2759 4/95/$4.00
ISBN: 0-7803-2759-4
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Theoretical
COUNTRY: United States
LANGUAGE: English
AB In many genetic **algorithm** applications the objective is to find a (near) optimal solution using a limited amount of computation. Given these requirements it is difficult to find a good balance between exploration and exploitation. Usually such a balance is found by tuning the various parameters (like the selective pressure, population size, the **mutation** and **crossover** rate) of the genetic **algorithm**. As an alternative we propose simultaneous tuning of the selective pressure and the **disruptiveness** of the **recombination** operators. Our experiments show that the combination of a proper selective pressure and a highly **disruptive recombination** operator yields superior performance. The reduction mechanism used in a steady state GA has a strong influence on the optimal **crossover disruptiveness**. Using the worst fitness deletion strategy the building blocks present in the current best individuals are always preserved. This releases the **crossover** operator from the burden to maintain good building blocks and allows us to tune **crossover disruptiveness** to improve the search for better individuals.

L105 ANSWER 17 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1994:4764248 INSPEC
DOCUMENT NUMBER: C9410-1250-220
TITLE: Automated discovery of detectors and iteration-performing calculations to recognize patterns in **protein sequences** using genetic programming.
AUTHOR: Koza, J.R. (Dept. of Comput. Sci., Stanford Univ., CA, USA)
SOURCE: Proceedings 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.94CH3405-8)
Los Alamitos, CA, USA: IEEE Comput. Soc. Press, 1994. p.684-9 of xvi+1009 pp. 12 refs.
Conference: Seattle, WA, USA, 21-23 June 1994
Sponsor(s): IEEE Comput. Soc. Tech. Committee on Pattern Anal. & Machine Intelligence
Price: CCCC 1063-6919/94/$3.00
ISBN: 0-8186-5825-8
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Application; Practical
COUNTRY: United States
LANGUAGE: English
AB This paper describes an automated process for the dynamic creation of a pattern-recognizing computer program consisting of initially unknown detectors, an initially-unknown iterative calculation incorporating the as-yet-uncreated detectors, and an initially-unspecified final calculation incorporating the results of the as-yet-uncreated iteration. The program's goal is to recognize a given **protein** segment as being a transmembrane domain or non-transmembrane area. The recognizing program to

solve this problem will be evolved using the recently developed genetic programming paradigm. Genetic programming starts with a primordial ooze of randomly generated computer programs composed of available programmatic ingredients and then genetically breeds the population using the Darwinian principle of survival of the fittest and the genetic **crossover** (sexual **recombination**) operation. Automatic function definition enables genetic programming to dynamically create subroutines (detectors). When cross-validated, the best genetically-evolved recognizer achieves an out-of-sample correlation of 0.968 and an out-of-sample error rate of 1.6%. This error rate is better than that recently reported for five other methods.

L105 ANSWER 18 OF 33 INSPEC (C) 2004 IEE on STN
ACCESSION NUMBER: 1992:4163718 INSPEC
DOCUMENT NUMBER: C9207-1180-009
TITLE: A formal analysis of the role of multi-point **crossover** in genetic **algorithms**.
AUTHOR: De Jong, K.A. (Dept. of Comput. Sci., George Mason Univ., Fairfax, VA, USA); Spears, W.M.
SOURCE: Annals of Mathematics and Artificial Intelligence (April 1992) vol.5, no.1, p.1-26. 11 refs.
CODEN: AMAIEC ISSN: 1012-2443
DOCUMENT TYPE: Journal
TREATMENT CODE: Theoretical
COUNTRY: Switzerland
LANGUAGE: English
AB Extends existing theoretical results in an attempt to provide a broader explanatory and predictive theory of the role of multi-point **crossover** in genetic **algorithms**. In particular, the authors extend the traditional **disruption** analysis to include two general forms of multi-point **crossover**: n-point **crossover** and uniform **crossover**. The authors also analyze two other aspects of multi-point **crossover** operators, namely their **recombination** potential and exploratory power. The results of this analysis provide a much clearer view of the role of multi-point **crossover** in genetic **algorithms**. The implications of these results on implementation issues and performance are discussed, and several directions for further research are suggested.

L105 ANSWER 19 OF 33 PASCAL COPYRIGHT 2004 INIST-CNRS. ALL RIGHTS RESERVED. on STN
ACCESSION NUMBER: 1995-0579821 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 1995 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): The Friedreich ataxia critical region spans a 150-kb interval on chromosome 9q13
AUTHOR: MONTERMINI L.; RODIUS F.; PIANESE L.; MOLTO M. D.; COSSEE M.; CAMPUZANO V.; CAVALCANTI F.; MONTICELLI A.; PALAU F.; GYAPAY G.; WENHERT M.; ZARA F.; PATEL P. I.; COCAZZA S.; KOENIG M.; PANDOLFO M.
CORPORATE SOURCE: Baylor coll. medicine, dep. neurology, Houston TX 77030, United States; INSERM CNRS, inst. molecular cellular biology (IGBMC), Strasbourg, France; Cent. hosp. univ., Strasbourg, France
SOURCE: American journal of human genetics, (1995), 57(5), 1061-1067, 17 refs.
ISSN: 0002-9297 CODEN: AJHGAG
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United States
LANGUAGE: English
AVAILABILITY: INIST-2610, 354000050534010110
AB By analysis of crossovers in key recombinant families and by homozygosity

analysis of inbred families, the Friedreich ataxia (FRDA) locus was localized in a 300-kb interval between the X104 gene and the microsatellite marker FR8 (D9S888). By homology searches of the **sequence databases**, we identified X104 as the human tight junction **protein** ZO-2 gene. We generated a large-scale physical map of the FRDA region by pulsed-field gel electrophoresis analysis of genomic DNA and of three YAC clones derived from different libraries, and we constructed an uninterrupted cosmid contig spanning the FRDA locus. The cAMP-dependent **protein** kinase y-catalytic subunit gene was identified within the critical FRDA interval, but it was excluded as candidate because of its biological properties and because of lack of **mutations** in FRDA patients. Six new polymorphic markers were isolated between FR2 (D9S886) and FR8 (D9S888), which were used for homozygosity analysis in a family in which parents of an affected child are distantly related. An ancient **recombination** involving the centromeric FRDA flanking markers had been previously demonstrated in this family. Homozygosity analysis indicated that the FRDA gene is localized in the telomeric 150 kb of the FR2-FR8 interval.

L105 ANSWER 20 OF 33 LIFESCI COPYRIGHT 2004 CSA on STN DUPLICATE 2
ACCESSION NUMBER: 2003:66956 LIFESCI
TITLE: In silico **Protein Recombination:** Enhancing Template and **Sequence** Alignment Selection for Comparative **Protein** Modelling
AUTHOR: Contreras-Moreira, B.; Fitzjohn, P.W.; Bates, P.A.
CORPORATE SOURCE: Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
SOURCE: Journal of Molecular Biology [J. Mol. Biol.], (20030502) vol. 328, no. 3, pp. 593-608. ISSN: 0022-2836.
DOCUMENT TYPE: Journal
FILE SEGMENT: N
LANGUAGE: English
SUMMARY LANGUAGE: English
AB Comparative modelling of **proteins** is a predictive technique to build an atomic model for a given amino acid **sequence**, on the basis of the structures of other **proteins** (templates) that have been determined experimentally. Critical problems arise in this procedure: selecting the correct templates, aligning the query **sequence** with them and building the non-conserved surface loops. In this work, we apply a genetic **algorithm**, with **crossover** and **mutation**, as a new tool to overcome the first two. In silico **protein recombination** proves to be an effective way to exploit the variability of templates and **sequence** alignments to produce populations of optimized models by artificial selection. Despite some limitations, the procedure is shown to be robust to alignment errors, while simplifying the task of selecting templates, making it a good candidate for automatic building of reliable **protein** models.

L105 ANSWER 21 OF 33 LIFESCI COPYRIGHT 2004 CSA on STN DUPLICATE 6
ACCESSION NUMBER: 97:116194 LIFESCI
TITLE: Tn3 resolvase catalyses multiple **recombination** events without intermediate rejoining of DNA ends
AUTHOR: McIlwraith, M.J.; Boocock, M.R.; Stark, W.M.*
CORPORATE SOURCE: Inst. Biomed. and Life Sci., Univ. Glasgow, 56 Dumbarton Rd., Glasgow G11 6NU, Scotland, UK
SOURCE: J. MOL. BIOL., (19970000) vol. 266, no. 1, pp. 108-121. ISSN: 0022-2836.
DOCUMENT TYPE: Journal
FILE SEGMENT: N
LANGUAGE: English
SUMMARY LANGUAGE: English

AB Resolvases and DNA invertases catalyse site-specific **recombination** by a concerted cut-and-religate mechanism. Topological data strongly suggest a rotational movement of the DNA half-sites during **recombination**: in an "iterative" mode of reaction, after cleavage of all four strands of the two recombining sites, the recombinase-linked half-sites seem to rotate through multiple steps of 180 degree prior to final religation. However, current structural data provide no clear support for the postulated corresponding rotation of enzyme subunits within an active tetramer. A key issue is whether repetition of apparent 180 degree rotation steps requires rejoining of the DNA strands and resetting of the catalytic machinery, or if multiple rotation steps can take place in the fully cleaved intermediate. We present evidence that a resolvase-catalysed DNA knotting reaction, brought about by apparent 360 degree rotation, can proceed without rejoining of the DNA strands in the recombinant (180 degree rotation) configuration. This behaviour is not compatible with a mechanism requiring a fixed arrangement of the catalytic subunits, and strongly suggests that **recombination** is **coupled** to **disruption** of the dimer interface between two subunits bound at each **crossover** site. We also show that an artificial supercoiled plasmid containing two res sites, with a single mismatched base-pair in one of the **crossover** sites, is a substrate for "suicidal" reactions in which resolvase remains covalently linked to two half-sites.

L105 ANSWER 22 OF 33 LIFESCI COPYRIGHT 2004 CSA on STN
ACCESSION NUMBER: 1999:20357 LIFESCI
TITLE: Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis
AUTHOR: Reyrat, J.-M.; Pelicic, V.; Gicquel, B.; Rappuoli, R.
CORPORATE SOURCE: IRIS Chiron-Vaccines, Via Fiorentina 1, 53100 Siena, Italy; E-mail: reyrat@iris02.biocine.it
SOURCE: Infection and Immunity, (19980900) vol. 66, no. 9, pp. 4011-4017.
ISSN: 0019-9567.
DOCUMENT TYPE: Journal
TREATMENT CODE: General Review
FILE SEGMENT: G; J; W3
LANGUAGE: English

AB The construction of clean and unmarked **mutations** in bacteria, where a gene is replaced by an in vitro-modified allele is a fundamental approach to the understanding of pathogenicity at a molecular level, the definition of structure-function relationships, and the production of vaccine candidates. With the increasing availability of complete bacterial genome **sequences**, the potential for such **mutagenesis** has grown exponentially. However, to date a great number of open reading frames (ORFs) remain unannotated since they present no homology with **sequences** already present in the **databases**. The precise function of some of these unknown genes will probably be deduced through in silico predictions, by comparing different genomes, or by the use of modern genetic strategies such as serial analysis of gene expression, in vivo expression technology, representational difference analysis, and signature tagged **mutagenesis** that help in extraction of information without a priori knowledge of the **sequence**. Nevertheless, reverse genetic analysis, one of the logical approaches to be undertaken, will be necessary to identify a phenotype and to attribute a precise function to many undefined ORFs. Reverse genetics is a powerful approach for the identification of gene function, in which the gene of interest is **mutated** or inactivated to study the resulting effects on the microorganism. Although allelic exchange is easy to perform with many bacteria, it remains very difficult or impractical with others. The classical method of using a suicide plasmid that is unable to replicate in the studied strain to deliver an inactivated allele of the gene in the chromosome is often not efficient because the frequency of

double **crossover** events may be low and because illegitimate **recombination** may occur. Consequently, allelic exchange **mutants** may represent only a small fraction of the transformants and may be difficult to isolate. Counterselectable markers are often instrumental for the construction of such **mutants**, especially in microorganisms for which the genetics schemes is poorly developed. Under appropriate growth conditions, a counterselectable gene promotes the death of the microorganisms harboring it. Hence, transformants which have integrated a suicide vector containing a counterselectable marker, either by a single event of homologous or illegitimate **recombination**, retain a copy of the counterselectable marker in the chromosome and are therefore eliminated in the presence of the counterselective compound. Consequently, counterselectable markers have been used for the positive selection of **mutants** that have undergone defined genetic alterations leading to the loss of the marker. In different studies, applications such as the construction of **mutants**, the isolation of insertion **sequence** (IS) elements, and the curing of plasmids have been described. The most-used counterselectable markers are the genes that confer sucrose, streptomycin, or fusaric acid sensitivity. They have been used to construct **mutants** or vaccine strains in Mycobacterium tuberculosis, Helicobacter pylori, Bordetella pertussis, and many other bacteria. Here we provide a short review of the situations in which the use of a counterselectable marker has proven to be particularly advantageous.

L105 ANSWER 23 OF 33 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
DUPLICATE 4

ACCESSION NUMBER: 1999:207922 BIOSIS
DOCUMENT NUMBER: PREV199900207922
TITLE: Identification of three aspartic acid residues essential for catalysis by the RusA Holliday junction resolvase.
AUTHOR(S): Bolt, Edward L.; Sharples, Gary J.; Lloyd, Robert G. [Reprint author]
CORPORATE SOURCE: Inst. Genetics, Univ. Nottingham, Queen's Med. Centre, Nottingham NG7 2UH, UK
SOURCE: Journal of Molecular Biology, (Feb. 19, 1999) Vol. 286, No. 2, pp. 403-415. print.
CODEN: JMOBAK. ISSN: 0022-2836.
DOCUMENT TYPE: Article
LANGUAGE: English
ENTRY DATE: Entered STN: 26 May 1999
Last Updated on STN: 26 May 1999

AB RusA is a Holliday junction resolvase encoded by the cryptic prophage DLP12 of Escherichia coli K-12 that can be activated to promote homologous **recombination** and DNA repair in resolution-deficient **mutants** lacking the RuvABC **proteins**. **Database** searches with the 120 amino acid residue RusA **sequence** identified 11 homologues from diverse species, including one from the extreme thermophile Aquifex aeolicus, which suggests that RusA may be of ancient bacterial ancestry. A multiple alignment of these **sequences** revealed seven conserved or invariant acidic residues in the C-terminal half of the E. coli **protein**. By making site-directed **mutations** at these positions and analyzing the ability of the **mutant proteins** to promote DNA repair in vivo and to resolve junctions in vitro, we identified three aspartic acid residues (D70, D72 and D91) that are essential for catalysis and that provide the first insight into the active-site mechanism of junction resolution by RusA. Substitution of any one of these three residues with asparagine reduces resolution activity >80-fold. The **mutant proteins** retain the ability to bind junction DNA regardless of the DNA **sequence** or of the mobility of the **crossover**. They interfere with the function of the RuvABC **proteins** in vivo, when expressed from a multicopy plasmid, an effect that is reproducible in

vitro and that reflects the fact that the RusA **proteins** have a higher affinity for junction DNA in the presence of Mg2+ than do the RuvA and RuvC **proteins**. The D70N **protein** has a greater affinity for junctions in Mg2+ than does the wild-type, which indicates that the negatively charged carboxyl group of the aspartate residue plays a critical role at the active site of RusA. Electrostatic repulsions between D70, D72 and D91 may help to form a classical Mg2+-binding pocket.

L105 ANSWER 24 OF 33 BIOSIS COPYRIGHT 2004 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 2002:212389 BIOSIS
DOCUMENT NUMBER: PREV200200212389
TITLE: Identification and characterization of the arcA gene of Shewanella oneidensis MR-1.
AUTHOR(S): Lies, D. P. [Reprint author]; Nealson, K. H.
CORPORATE SOURCE: California Institute of Technology, Pasadena, CA, USA
SOURCE: Abstracts of the General Meeting of the American Society for Microbiology, (2001) Vol. 101, pp. 401. print.
Meeting Info.: 101st General Meeting of the American Society for Microbiology. Orlando, FL, USA. May 20-24, 2001. American Society of Microbiology.
ISSN: 1060-2011.
DOCUMENT TYPE: Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)
LANGUAGE: English
ENTRY DATE: Entered STN: 27 Mar 2002
Last Updated on STN: 27 Mar 2002

AB Shewanella oneidensis MR-1 is an obligatorily respiratory organism capable of growth with either oxygen or any of at least 12 other alternate electron acceptors in the absence of oxygen. We have identified, cloned and **mutated** the gene encoding ArcA, one of the primary regulators of the oxic to anoxic growth transition in Escherichia coli, from S. oneidensis MR-1. ArcA serves primarily as a negative regulator of aerobic metabolism in E. coli but also is involved in positive regulation of genes expressed under microaerobic conditions and some anaerobically expressed genes. The E. coli ArcA **protein sequence** was used to search the unfinished S. oneidensis MR-1 genome **sequence database** available from. TIGR. The MR-1 ArcA **protein** was 72-81% identical to homologs from E. coli, V. cholerae and H. influenzae, including the conserved phosphate-accepting aspartate residue at amino acid 54 in E. coli. The MR-1 arcA DNA **sequence** revealed that open reading frames both upstream and downstream of arcA are transcribed in the opposite direction from arcA, indicating that it is not part of an operon. The arcA gene is also followed by **sequences** typical of Rho-independent transcriptional terminators. No obvious binding sites for either Fnr (TTGAT-ATCAA) or ArcA ((A/T)GTTAATTA(A/T)) **proteins** were identified upstream of the MR-1 arcA gene but the gene was preceded by a tandem repeat of TGGTTA(G/A)AAT(A/T)T. No significant homolog of the ArcB **protein** has been detected in the MR-1 genome. The MR-1 arcA gene was cloned and **mutated** using **crossover** PCR to delete all except the first and last seven codons of the gene, separated by a 21 bp linker region encoding a unique SwaI restricton site. This **mutated** gene, and a version containing a kanamycin resistance cassette cloned into the SwaI site, was introduced into the MR-1 genome by homologous **recombination** to replace the wild-type allele. These **mutants** are being analyzed for the effects of the arcA **mutations** on aerobic, microaerobic, and anaerobic growth of MR-1, as well as synthesis of enzymes such as pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase that are known to be repressed during anaerobic growth of MR-1.

L105 ANSWER 25 OF 33 ANABSTR COPYRIGHT 2004 RSC on STN DUPLICATE 5
AB A genetic **algorithm** (GA) was compared to a simulated annealing method for the elucidation of the structure of the natural heptapeptide

optoid .mu.-selective dermorphin. The molecule was represented by its 31 internal co-ordinates (torsion angles), which had to be optimized so as to minimize atomic overlap and distance-restraint violations. In this instance the **crossover** operator of the GA was not able to contribute significantly to its performance, which was therefore dependant mostly on selection and point **mutation**. As a result, the more sophisticated **mutation**/selection scheme imposed by simulated annealing outperformed the GA in structure elucidation. The GA found conformations of similar quality to those found by simulated annealing, but took three times as long to converge on the solutions. The theories of GA and simulated annealing are briefly described and the results are discussed.

L105 ANSWER 26 OF 33 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN
ACCESSION NUMBER: 2001-536358 [59] WPIDS
DOC. NO. CPI: C2001-159641
TITLE: Device or integrated system for generating diverse **nucleic acids** comprises an array of shuffled or mutagenized **nucleic acids**, or transcribed versions, and in vitro translation reagents.
DERWENT CLASS: B04 D16
INVENTOR(S): AFFHOLTER, J A; BASS, S H; CARR, B; CRAMERI, A; DAVIS, S C; EMIG, R; GIVER, L J; GOLDMAN, S; GUSTAFSSON, C; JENNE, S; LONGCHAMP, P; MINSHULL, J; PATTEN, P A; RAILLARD, S A; STEMMER, W P C; TOBIN, M; WELCH, M; LONGSCHAMP, P
PATENT ASSIGNEE(S): (MAXY-N) MAXYGEN INC
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001051663	A2	20010719	(200159)*	EN	258
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW					
AU 2001027881	A	20010724	(200166)		
US 2001039014	A1	20011108	(200171)		
EP 1276900	A2	20030122	(200308)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR					
US 2003054383	A1	20030320	(200323)		
US 2003054384	A1	20030320	(200323)		
US 2003064393	A1	20030403	(200325)		
JP 2003519495	W	20030624	(200341)		283

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2001051663	A2		WO 2001-US1056	20010110
AU 2001027881	A		AU 2001-27881	20010110
US 2001039014	A1	Provisional	US 2000-175551P	20000111
		Provisional	US 2000-213947P	20000623
			US 2001-760010	20010110
EP 1276900	A2		EP 2001-902036	20010110
			WO 2001-US1056	20010110
US 2003054383	A1	Provisional	US 2000-175551P	20000111
		Provisional	US 2000-213947P	20000623
		Cont of	US 2001-760010	20010110

		US 2002-154936	20020523
US 2003054384 A1	Provisional	US 2000-175551P	20000111
	Provisional	US 2000-213947P	20000623
	Cont of	US 2001-760010	20010110
		US 2002-154939	20020523
US 2003064393 A1	Provisional	US 2000-175551P	20000111
	Provisional	US 2000-213947P	20000623
	Cont of	US 2001-760010	20010110
		US 2002-155739	20020523
JP 2003519495 W		JP 2001-551237	20010110
		WO 2001-US1056	20010110

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001027881 A		Based on	WO 2001051663
EP 1276900	A2	Based on	WO 2001051663
JP 2003519495 W		Based on	WO 2001051663

PRIORITY APPLN. INFO: US 2000-213947P 20000623; US 2000-175551P 20000111; US 2001-760010 20010110; US 2002-154936 20020523; US 2002-154939 20020523; US 2002-155739 20020523

AB WO 200151663 A UPAB: 20030919

NOVELTY - A device or integrated system (I), comprising:

(1) a physical or logical array of reaction mixtures (II), each reaction mixture comprising shuffled or mutagenized **nucleic acids** (SNAs or MNAs) or transcribed SNAs or MNAs; and

in vitro translation reagents (IVTLR), is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) a diversity generation device (III) comprising a programmed thermocycler and a fragmentation module coupled to the thermocycler;

(2) a diversity generation kit (IV) comprising (III) and one more reagents for diversity generation;

(3) a method (M1) of processing (SNAs) or (MNAs) comprising:

(a) providing (II), in which a number of reaction mixtures (RM) comprise members of a first population of **nucleic acids** (Na1) comprising (SNAs), or transcribed (SNAs), or (MNAs) or transcribed (MNAs), where many of the (RM) further comprise an in vitro translation reactant (IVTLR); and

(b) detecting in vitro translation products by the members of (II);

(4) the physical or logical array of reaction mixtures (II) produced by (M1);

(5) a method (M2) of recombining members of a physical or logical array of **nucleic acids** (V) comprising:

(a) providing (Na1); or

(b) providing a **data structure** comprising character strings corresponding to (Na1);

(c) recombining members of (Na1), providing a first population of recombinant **nucleic acids** (rNa1); or

(d) recombining the character strings corresponding to members of (Na1), providing a population of character strings corresponding to (rNa1), and into (rNa1);

(e) spatially or logically separating members of (rNa1) to produce (V) and amplifying the recombinant **nucleic acids** in (V) in vitro ; or

in vitro amplifying members of (rNa1) and physically or logically separating them to produce an amplified (V);

(6) a method (M3) of recombining members of (V) comprising:

(a) providing (Na1) arranged in a physical or logical array;

(b) recombining members of (Na1) with additional **nucleic**

acids, providing (rNa1);

(c) amplifying (rNa1) in the physical or logical array; and

(d) screening the first or amplified (rNa1), or a duplicate, for a desired property;

(7) a method (M4) of detecting or enriching for in vitro transcription or translation products comprising:

(a) localizing first **nucleic acids** which encode moieties proximal to moiety recognition agents which specifically bind to them;

in vitro translating or transcribing the **nucleic acids**, producing moieties which diffuse or flow into contact with the recognition agents; and

(b) permitting binding of the moieties to the recognition agents, and detecting or enriching for moieties by detecting or collecting material proximal to, within or contiguous with the moiety recognition agent;

(8) a solid substrate (VI) made by (M4);

(9) a method (M5) of producing duplicate arrays of (SNAs) or (MNAs) comprising:

(a) providing (II); and

(b) forming a duplicate array of copies of (II) by physically or logically organizing the copies into a physical or logical array;

(10) the physical or logical array and duplicate array (VII) produced by (M5);

(11) a method (M6) of normalizing an array of reaction mixtures comprising:

in vitro transcribing or translating (II) to produce an array of products; and

(a) determining a correction factor which accounts for variation in concentration of the products at different sites in the array of products;

(12) the physical or logical array of (SNAs) or (MNAs) or transcribed (SNAs) or (MNAs), the array of products and the secondary array (VIII) produced by (M6);

(13) a method (M7) of recombining one or more nucleic acids comprising:

(a) immobilizing one or more template nucleic acids on a solid support;

(b) annealing overlapping complementary nucleic acid fragments to the immobilized template nucleic acid;

(c) extending or ligating the annealed fragments to produce at least one heteroduplex, which comprises a template nucleic acid and a substantially full-length heterolog complementary to the template nucleic acid; and

(d) recovering at least one substantially full-length heterolog;

(14) a full-length heterolog (IX) produced by (M7)

(15) an array (X) comprising heteroduplexes or full-length heterologs produced by (M7);

(16) an amplified heterolog (X) produced by (M7);

(17) a vector (XII) produced by (M7);

(18) a cell (XIII) produced by (M7);

(19) a library of diversified heterologs (XIV) produced by (M7);

(20) an integrated system (XV) comprising an array which comprises heteroduplexes or full-length heterologs produced by (M7);

(21) a method (M8) of directing nucleic acid fragmentation using a computer, the method comprising calculating the ratio of uracil to thymidine which may then be used in a fragmentation module to produce one or more nucleic acid fragments of a selected length;

(22) a method (M9) of directing polymerase chain reaction (PCR) using a computer, the method comprising calculating one or more crossover regions between two or more parental nucleic acid sequences using one or more annealing temperature or extension temperature;

(23) a method (M10) of selecting one or more parental nucleic acids for diversity generation using a computer comprising:

(a) performing an alignment between two or more potential parental

nucleic acid sequences;
(b) calculating a number of mismatches between alignment;
(c) calculating a melting temperature for one ore more window of w bases in the alignment;
(d) identifying one or more window of w bases having a melting temperature greater than x;
(e) identifying one or more crossover segment in the alignment which comprises two or more windows having a melting temperature greater than x and that are separated by no more than n nucleotides;
(f) calculating a dispersion of the one or more crossover segments;
(g) calculating a first score for each alignment based on the number of windows having a melting temperature greater than x, the dispersion, and the number of crossover segments identified;
(h) calculating a second score based on the number of mismatches, the number of windows having a melting temperature greater than x, the dispersion, and the number of crossover segments identified; and
(i) selecting one or more parental nucleic acid based on the first score and/or the second score; and
(24) a web page (XVI) for directing nucleic acid diversity generation comprising a computer readable medium that causes a computer to perform (M8), (M9) or (M10).
USE - (I) is useful for performing nucleic acid recombination, mutation, shuffling, and other diversity generating reactions in vitro to generate diverse nucleic acids and screen for desirable properties of those nucleic acids such as their products, e.g., encoded RNAs (catalytic RNAs, ribozymes) or proteins.
ADVANTAGE - Each aspect of diversity generation and downstream screening processes can be automated and used individually in separate modules or collectively in an integrated system or device.
Dwg.0/33

L105 ANSWER 27 OF 33 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN
ACCESSION NUMBER: 2000-491075 [43] WPIDS
CROSS REFERENCE: 2000-303449 [26]; 2000-482862 [42]; 2000-514667 [46]; 2001-300096 [31]; 2003-777161 [73]
DOC. NO. NON-CPI: N2000-364428
DOC. NO. CPI: C2000-147612
TITLE: In silico recombinant **nucleic acid** preparation by genetic **algorithm** guided gene synthesis involves providing a number of parental character strings, providing **oligonucleotides** and elongating them.
DERWENT CLASS: B04 D16 T01
INVENTOR(S): DEL CARDAYRE, S; GUSTAFSSON, C; MINSHULL, J; PATTEN, P A; SELIFONOV, S A; STEMMER, W P C; TOBIN, M
PATENT ASSIGNEE(S): (MAXY-N) MAXYGEN INC
COUNTRY COUNT: 91
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2000042560	A2	20000720	(200043)*	EN	127
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW					
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW					
AU 2000032101	A	20000801	(200054)		
EP 1062614	A1	20001227	(200102)	EN	
R: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE					
KR 2001042037	A	20010525	(200168)		
JP 2002534965	W	20021022	(200301)		161

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2000042560	A2	WO 2000-US1202	20000118
AU 2000032101	A	AU 2000-32101	20000118
EP 1062614	A1	EP 2000-909922	20000118
		WO 2000-US1202	20000118
KR 2001042037	A	KR 2000-710375	20000919
JP 2002534965	W	JP 2000-594067	20000118
		WO 2000-US1202	20000118

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2000032101	A	Based on	WO 2000042560
EP 1062614	A1	Based on	WO 2000042560
JP 2002534965	W	Based on	WO 2000042560

PRIORITY APPLN. INFO: US 2000-416837 20000118; US 1999-116447P 19990119; US 1999-118813P 19990205; US 1999-118854P 19990205; US 1999-141049P 19990624; US 1999-408392 19990928; US 1999-408393 19990928; US 1999-416375 19991012; US 1999-416837 19991012; US 2000-416375 20000118

AB WO 200042560 A UPAB: 20031112

NOVELTY - Preparing recombinant **nucleic acid** (I) from **oligonucleotides** which correspond to a set of character string subsequences (SCSS) comprising at least two parental character strings (PCS) corresponding to a number of **nucleic acids**, is new.

DETAILED DESCRIPTION - Preparing recombinant **nucleic acid** (I) by aligning for maximum identity a number of parental character strings (PCS) corresponding to a number of **nucleic acids**, defining a set of character string subsequences (SCSS) comprising at least two of the PCS, providing a set of **oligonucleotides** corresponding to the SCSS and then annealing and elongating one or more **oligonucleotides** with polymerase or ligating at least two with ligase to produce (I).

INDEPENDENT CLAIMS are also included for the following:

(1) preparing character strings (CS) by providing PCS encoding a polynucleotide or **polypeptide**, providing a set of **oligonucleotide** character strings of preselected length that encode a number of single-stranded **oligonucleotide sequence** comprising **sequence** fragments of PCS and its complement and creating a set of derivatives of parental **sequence** comprising **sequence** variant strings, a set of multiple **mutations** with one **mutation** per variant string;

(2) a library prepared by the above said method;

(3) facilitating **recombination** between two or more divergent **nucleic acids** by aligning PCS corresponding to divergent **nucleic acids**, identifying regions of **sequence** identity and regions of **sequence** diversity, defining a diplomat CS which is intermediate in PCS, synthesizing at least a portion of the diplomat **sequence** to produce a diplomat **nucleic acid** and recombining a mixture of parental **nucleic acid** and diplomat **nucleic acid**
;

(4) a mixture of selected **nucleic acids** produced by the above said method;

(5) generating and recombining **nucleic acids** by inputting a number of amino acid **sequence** character strings (ASCS) into a digital system, reverse translating ASCS in the digital system to a number of **nucleic acid** character strings which are species codon biased in a selected expression host and with optimized **sequence** similarity between a number of **nucleic acid** character strings and synthesizing one or more **oligonucleotides** from one or more reverse translated **nucleic acid sequences;**

(6) optimizing activity of a **nucleic acid** by parameterizing a number of **nucleic acids** or **proteins** to provide a set of multidimensional datapoints, extrapolating one or more postulated multidimensional datapoint from the set of multidimensional datapoints and converting the postulated multidimensional datapoint to a new CS corresponding to a postulated **nucleic acid** or **protein;**

(7) providing a library of recombinant **nucleic acids** which is enriched for a **sequence** of interest and selecting the library by producing an initial library of at least about 106 recombinant **nucleic acids,** comprising at least about 105 different non-identical units, hybridizing the library to one or more population of **nucleic acids** that correspond to one or more subsequences in the different library units;

(8) the enriched library produced by the above said method;

(9) generating a library of biological **polymers** by generating a diverse population of CS in a computer, which in turn are generated by alteration of pre-existing CS, synthesizing the diverse population of CS in which diverse population comprises the library of biological **polymers;** and

(10) an integrated system comprising a computer having a first data set comprising a first CS, a second data set comprising a second CS, software for aligning the first and second CSs, software for performing a genetic operation on the first or second CS, an output file comprising a third data set comprising a third CS, the third CS comprising CS subsequences from the first and second CSs, and an **oligonucleotide sequence** output file comprising a plurality of overlapping **oligonucleotide sequences** corresponding to third CS.

USE - The method is useful for rapid evolution of **nucleic acids** in vitro and in vivo and provides for generation of encoded molecules with new and/or improved properties. **Proteins** and **nucleic acids** of industrial, agricultural and therapeutic importance can be created or improved through DNA shuffling procedures.

ADVANTAGE - Physical access to genes or organisms is not required as **sequence** information is used for design and selection of oligo. Extensive **sequence** information is provided and **sequences** from inaccessible, non cultivable organisms can also be used. **Sequences** from pathogens without actual handling of pathogens and all type **sequences** including damaged and incomplete genes are amenable to this method. All genetic operators and crossovers can be fully and independently controlled in a reproducible fashion removing human error and variability from physical experiments with DNA manipulations. **Sequences** with frame-shift **mutations** are eliminated or fixed. Wild type parents do not contaminate derivative libraries with multiple redundant parental molecules.

Dwg.0/15

L105 ANSWER 28 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 2003:273057 SCISEARCH
THE GENUINE ARTICLE: 657WL
TITLE: Meiotic chromosome synapsis and **recombination** in Arabidopsis thaliana; an integration of cytological and molecular approaches

AUTHOR: Jones G H (Reprint); Armstrong S J; Caryl A P; Franklin F C H
CORPORATE SOURCE: Univ Birmingham, Sch Biosci, Birmingham B15 2TT, W Midlands, England (Reprint)
COUNTRY OF AUTHOR: England
SOURCE: CHROMOSOME RESEARCH, (MAR 2003) Vol. 11, No. 3, pp. 205-215.
Publisher: KLUWER ACADEMIC PUBL, VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS.
ISSN: 0967-3849.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 54

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Arabidopsis has emerged as an important model for the analysis of meiosis in Angiosperm plants, creating an interesting and useful parallel to other model organisms. This development has been underpinned by advances in the molecular biology and genetics of Arabidopsis, especially the determination of its entire genome **sequence**. However, these advances alone would have been insufficient without the development of improved methods for cytological analysis and cytogenetic investigation of meiotic nuclei and chromosomes. A basic descriptive framework of meiosis in Arabidopsis has been established based on these procedures. In addition, molecular cytogenetic and immunocytological techniques have provided supplementary detailed information on some aspects of meiosis. Gene identification and characterization have proceeded in parallel with these developments based on both forward and reverse genetic procedures utilising the considerable range of Arabidopsis genetic and molecular resources, such as T-DNA and transposon tagged lines as well as the genomic DNA **database**, in combination with cytological analysis. A diverse range of meiotic genes have been identified and analysed by these procedures and in selected cases they have been subjected to detailed functional analysis. This review focuses on genes that are involved in the key meiotic events of chromosome synapsis and **recombination**.

L105 ANSWER 29 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 2000:516262 SCISEARCH
THE GENUINE ARTICLE: 329ZN
TITLE: Integration and excision of a Bacteroides conjugative transposon, CTnDOT
AUTHOR: Cheng Q I; Paszkiet B J; Shoemaker N B; Gardner J F; Salyers A A (Reprint)
CORPORATE SOURCE: UNIV ILLINOIS, DEPT MICROBIOL, 601 S GOODWIN AVE, URBANA, IL 61801 (Reprint); UNIV ILLINOIS, DEPT MICROBIOL, URBANA, IL 61801
COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF BACTERIOLOGY, (JUL 2000) Vol. 182, No. 14, pp. 4035-4043.
Publisher: AMER SOC MICROBIOLOGY, 1752 N ST NW, WASHINGTON, DC 20036-2904.
ISSN: 0021-9193.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 34

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Bacteroides conjugative transposons (CTns) are thought to transfer by first excising themselves from the chromosome to form a nonreplicating circle, which is then transferred by conjugation to a recipient. Earlier studies showed that transfer of most Bacteroides CTns is stimulated by tetracycline, but it was not known which step in transfer is regulated. We have cloned and sequenced both ends of the Bacteroides CTn, CTnDOT, and

have used this information to examine excision and integration events. A segment of DNA that contains the joined ends of CTnDOT and an adjacent open reading frame (ORF), intDOT, was necessary and sufficient for integration into the Bacteroides chromosome. Integration of this miniature form of the CTn was not regulated by tetracycline. Excision of CTnDOT and formation of the circular intermediate were detected by PCR, using primers designed from the end sequences. Sequence analysis of the PCR products revealed that excision and integration involve a 5-bp **coupling** sequence-type mechanism possibly similar to that used by CTn Tn916, a CTn found originally in enterococci. PCR analysis also demonstrated that excision is a tetracycline-regulated step in transfer. The integrated minielement containing intDOT and the ends of CTnDOT did not excise, nor did a larger minielement that also contained an ORF located immediately dowmstream of intDOT designated orf2. Thus, excision involves other genes besides intDOT and orf2. Both intDOT and orf2 were **disrupted** by single-**crossover** insertions. Analysis of the **disruption** mutants showed that intDOT was essential for excision but orf2 was not. Despite its proximity to the integrase gene, orf2 appears not to be essential for excision.

L105 ANSWER 30 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 2000:545928 SCISEARCH
THE GENUINE ARTICLE: 333LY
TITLE: Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells
AUTHOR: Johnson R D; Jasin M (Reprint)
CORPORATE SOURCE: MEM SLOAN KETTERING CANC CTR, CELL BIOL PROGRAM, 1275 YORK AVE, NEW YORK, NY 10021 (Reprint); MEM SLOAN KETTERING CANC CTR, CELL BIOL PROGRAM, NEW YORK, NY 10021; CORNELL UNIV, GRAD SCH MED SCI, NEW YORK, NY 10021
COUNTRY OF AUTHOR: USA
SOURCE: EMBO JOURNAL, (3 JUL 2000) Vol. 19, No. 13, pp. 3398-3407. Publisher: OXFORD UNIV PRESS, GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND. ISSN: 0261-4189.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 54

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB In mammalian cells, repair of DNA double-strand breaks (DSBs) occurs by both homologous and nonhomologous mechanisms. By definition, homologous **recombination** requires a template with sufficient sequence identity to the damaged molecule in order to direct repair, We now show that the sister chromatid acts as a repair template in a substantial proportion of DSB repair events. The outcome of sister chromatid repair is primarily gene conversion unassociated with reciprocal exchange. This contrasts with expectations from the classical DSB repair model originally proposed for yeast meiotic **recombination**, but is consistent with models in which **recombination** is **coupled** intimately with replication. These results may explain why cytologically observable sister chromatid exchanges are induced only weakly by DNA-damaging agents that cause strand breaks, since most homologous repair events would not be observed. A preference for non-**crossover** events between sister chromatids suggests that crossovers, although genetically silent, mag be disfavored for other reasons. possibly, a general bias against **crossing over** in mitotic cells exists to reduce the potential for genome alterations when other homologous repair templates are utilized.

L105 ANSWER 31 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 1998:53584 SCISEARCH
THE GENUINE ARTICLE: YP769

TITLE: The nature of polymorphism of the HLA class I non-coding regions and their contribution to the diversification of HLA
AUTHOR: Blasczyk R (Reprint); Kotsch K; Wehling J
CORPORATE SOURCE: HUMBOLDT UNIV BERLIN, DEPT INTERNAL MED, DIV HEMATOL & ONCOL, BLOODBANK, VIRCHOW KLINIKUM, D-13353 BERLIN, GERMANY (Reprint)
COUNTRY OF AUTHOR: GERMANY
SOURCE: HEREDITAS, (OCT 1997) Vol. 127, No. 1-2, pp. 7-9.
Publisher: HEREDITAS-DISTRIBUTION, GJORLOFFSGATAN 121, 261 34 LANDSKRONA, SWEDEN.
ISSN: 0018-0661.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 10
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The **sequence database** of HLA class I genes is mainly derived from mRNA analysis. Little is known about the non-coding **sequences** of the different class I alleles. In this study we have determined the **sequence** of the Ist through 3rd introns of the majority of HLA-A and -B alleles. The few published **sequences** emerged to contain substantial errors. The introns turned out to be highly polymorphic with a variability of 14.6% in the Ist intron decreasing to 6.2% in the 3rd intron. Against all expectations, this variability is not characterised by random point **mutations** but by a highly systematic diversity reflecting the ancestral relationship of the HLA alleles. The variability is arrested on the level of the serological diversity. The striking conservation within each ancestral lineage suggests that point **mutations** have been negatively selected. This finding could be explained by the evolutionary pressure on base order, promoting the potential to extrude single-strand stem-loops from supercoiled duplex DNA, which is believed to be important for **recombination**. Moreover, the GC content was found to be as high as 78% in the ist and 2nd introns and 55% in the 3rd intron. These CpG islands are directly involved in the exchange of short stretches of DNA in unequal **crossing-over** events. Additionally, conversion between different class I **sequences** is Facilitated by regions of strong homology, stabilizing the pairing of variable regions. All these observations indicate the potential of a substantial contribution of introns to the recombinational activity of class I genes. The exclusive clustering of CpG islands in the Ist and 2nd introns restricts the gene conversion events to the regions of the 2nd and 3rd exons and therefore protects the conservation of the 5' flanking region and the 3' part of the gene. Since there are less diversification forces acting on introns they may be more conserved in a trans-species manner than exons. Therefore, they could provide the answer for the controversy regarding intra-or trans-species evolution.

L105 ANSWER 32 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 95:789652 SCISEARCH
THE GENUINE ARTICLE: TD899
TITLE: A GENETIC-LINKAGE MAP FOR THE DOMESTICATED SILKWORM, BOMBYX-MORI, BASED ON RESTRICTION-FRAGMENT-LENGTH-POLYMORPHISMS
AUTHOR: SHI J R; HECKEL D G; GOLDSMITH M R (Reprint)
CORPORATE SOURCE: UNIV RHODE ISL, DEPT ZOOL, KINGSTON, RI, 02881 (Reprint); UNIV RHODE ISL, DEPT ZOOL, KINGSTON, RI, 02881; CLEMSON UNIV, DEPT BIOL SCI, CLEMSON, SC, 29634
COUNTRY OF AUTHOR: USA
SOURCE: GENETICAL RESEARCH, (OCT 1995) Vol. 66, No. 2, pp. 109-126

ISSN: 0016-6723.

DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: ENGLISH
REFERENCE COUNT: 98

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB We present data for the initial construction of a molecular linkage map for the domesticated silkworm, Bombyx mori, based on 52 progeny from an F2 cross from a pair mating of inbred strains p50 and C108, using restriction fragment length polymorphisms (RFLPs). The map contains 15 characterized single copy **sequences**, 36 anonymous **sequences** derived from a follicular cDNA library, and 10 loci corresponding to a low copy number retrotransposon, mag. The 15 linkage groups and 8 ungrouped loci account for 23 of the 28 chromosomes and span a total **recombination** length of 413 cM; 10 linkage groups were correlated with established classic genetic maps. Scoring data from Southern blots were analysed using two Pascal programs written specifically to analyse linkage data in Lepidoptera, where females are the heterogametic sex and have achiasmatic meiosis (no **crossing-over**). These first examine evidence for linkage by calculating the maximum lod score under the hypothesis that the two loci are linked over the likelihood under the hypothesis that the two loci assort independently, and then determine multilocus linkage maps for groups of putatively syntenic loci by calculating the maximum likelihood estimate of the **recombination** fractions and the log likelihood using the EM **algorithm** for a specified order of loci along the chromosome. In addition, the possibility of spurious linkage was exhaustively tested by searching for genotypes forbidden by the absence of **crossing-over** in one sex.

L105 ANSWER 33 OF 33 SCISEARCH COPYRIGHT 2004 THOMSON ISI on STN
ACCESSION NUMBER: 92:658719 SCISEARCH
THE GENUINE ARTICLE: JW888
TITLE: TARGETED GENE REPLACEMENTS IN A STREPTOMYCES POLYKETIDE SYNTHASE GENE-CLUSTER - ROLE FOR THE ACYL CARRIER PROTEIN
AUTHOR: KHOSLA C; EBERTKHOSLA S; HOPWOOD D A (Reprint)
CORPORATE SOURCE: JOHN INNES INST, DEPT GENET, JOHN INNES CTR, COLNEY LANE, NORWICH NR4 7UH, NORFOLK, ENGLAND
COUNTRY OF AUTHOR: ENGLAND
SOURCE: MOLECULAR MICROBIOLOGY, (NOV 1992) Vol. 6, No. 21, pp. 3237-3249.
ISSN: 0950-382X.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: ENGLISH
REFERENCE COUNT: 49

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB A methodology was developed to construct any desired chromosomal mutation in the gene cluster that encodes the actinorhodin polyketide synthase (PKS) of Streptomyces coelicolor A3(2). A positive selection marker (resistance gene) is first introduced by double **crossing-over** into the chromosomal site of interest by use of an unstable delivery plasmid. This marker is subsequently replaced by the desired mutant allele via a second high-frequency double **recombination** event. The technology has been used to: (i) explore the significance of translational **coupling** between two adjacent PKS genes; (ii) prove that the acyl carrier protein (ACP) encoded by a gene in the cluster is necessary for the function of the actinorhodin PKS; (iii) provide genetic evidence supporting the hypothesis that serine 42 is the site of phosphopantetheinylation in the ACP of the actinorhodin PKS; and (iv) demonstrate that this ACP can be replaced by a Saccharopolyspora fatty acid synthase ACP to generate an active hybrid PKS.

FILE 'HOME' ENTERED AT 16:22:53 ON 08 JAN 2004